



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

ISABELA MENDES BONFIM

MOLECULAR BASIS OF CARBOHYDRATE DEGRADATION AND
UTILIZATION SYSTEMS FROM THE PHYTOPATHOGEN
XANTHOMONAS CITRI PV. *CITRI*

ESTUDOS DAS BASES MOLECULARES DOS SISTEMAS DE
DEGRADAÇÃO E UTILIZAÇÃO DE CARBOIDRATOS DO
FITOPATÓGENO *XANTHOMONAS CITRI* PV. *CITRI*

CAMPINAS
2022

ISABELA MENDES BONFIM

MOLECULAR BASIS OF CARBOHYDRATE DEGRADATION AND
UTILIZATION SYSTEMS FROM THE PHYTOPATHOGEN
XANTHOMONAS CITRI PV. *CITRI*

ESTUDOS DAS BASES MOLECULARES DO SISTEMAS DE
DEGRADAÇÃO E UTILIZAÇÃO DE CARBOIDRATOS DO
FITOPATÓGENO *XANTHOMONAS CITRI* PV. *CITRI*

Thesis presented to the Biology
Institute of the University of Campinas
in partial fulfillment of the
requirements for the degree of Doctor
in Molecular and Morphofunctional
Biology, in the area of Biochemistry.

Tese apresentada ao Instituto de
Biologia da Universidade Estadual de
Campinas como parte dos requisitos
exigidos para a obtenção do título de
Doutora em Biologia Molecular e
Morfofuncional, na Área de
Bioquímica.

Orientador: DR. MÁRIO TYAGO MURAKAMI

Co-orientadora: DRA. PRISCILA OLIVEIRA DE GIUSEPPE

ESTE ARQUIVO DIGITAL CORRESPONDE À
VERSÃO FINAL DA TESE DEFENDIDA PELA
ALUNA ISABELA MENDES BONFIM, E
ORIENTADA PELO PROF. DR. MÁRIO TYAGO
MURAKAMI.

CAMPINAS
2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Gustavo Lebre de Marco - CRB 8/7977

B641m Bonfim, Isabela Mendes, 1995-
Molecular basis of carbohydrate degradation and utilization systems from the phytopathogen *Xanthomonas citri* pv. *citri* / Isabela Mendes Bonfim. – Campinas, SP : [s.n.], 2022.

Orientador: Mário Tyago Murakami.

Coorientador: Priscila Oliveira de Giuseppe.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Carboidratos. 2. Enzimas. 3. RNA-seq. 4. *Xanthomonas*. 5. Polissacarídeos. I. Murakami, Mário Tyago. II. Giuseppe, Priscila Oliveira de. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações Complementares

Título em outro idioma: Estudos das bases moleculares dos sistemas de degradação e utilização de carboidratos do fitopatógeno *Xanthomonas citri* pv. *citri*

Palavras-chave em inglês:

Carbohydrates

Enzymes

RNA-seq

Xanthomonas

Polysaccharides

Área de concentração: Bioquímica

Titulação: Doutora em Biologia Funcional e Molecular

Banca examinadora:

Mário Tyago Murakami

Cristina Elisa Alvarez Martinez

André Ricardo de Lima Damásio

Marcio Vinicius Bertacine Dias

Maxuel de Oliveira Andrade

Data de defesa: 07-11-2022

Programa de Pós-Graduação: Biologia Molecular e Morfofuncional

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-1957-2458>

- Currículo Lattes do autor: <http://lattes.cnpq.br/5843220450300990>

Campinas, 07 de novembro de 2022.

COMISSÃO EXAMINADORA

Prof. Dr. Mário Tyago Murakami

Profa. Dra. Cristina Elisa Alvarez Martinez

Prof. Dr. André Ricardo de Lima Damásio

Prof. Dr. Marcio Vinicius Bertacine Dias

Prof. Dr. Maxuel de Oliveira Andrade

Os membros da Comissão Examinadora acima assinaram a Ata de defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Pós-Graduação em Biologia Molecular e Morfofuncional da Unidade do Instituto de Biologia da Unicamp.

Dedicatória

Com carinho aos meus pais e primeiros
professores, Beto e Ângela

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Código de Financiamento 001) e da Fundação de Amparo à Pesquisa do estado de São Paulo (FAPESP; Processo 2017/00203-0 vinculado ao Projeto Temático Processo FAPESP 2015/26982-0).

Agradeço ao meu orientador Dr. Mário Murakami pela oportunidade, por todo apoio ao longo do projeto, pela confiança, paciência, por todas as discussões científicas e por sempre me incentivar a buscar novos desafios.

Agradeço a minha orientadora Dra. Priscila Giuseppe pelo grande apoio em cada etapa, pela paciência, sábios ensinamentos e todo carinho, proporcionando discussões enriquecedoras e sempre me incentivando a ser corajosa e resiliente frente a todos os desafios dessa jornada.

À Dra. Gabriela Persinoti, Douglas Paixão e Joaquim Martins por todo apoio e discussões científicas realizadas ao longo do planejamento experimental, análise de dados de transcriptômica e filogenia que foram essenciais para o desenvolvimento e concepção desse projeto.

À Dra. Cristina Elisa Alvarez Martinez e ao Dr. André Ricardo de Lima Damásio, Dr. Marcio Vinicius Bertacine Dias e Dr. Maxuel de Oliveira Andrade pela participação na banca examinadora desse trabalho.

À Marcele pela enorme parceria ao longo do doutorado, por todas as contribuições científicas, carinho e amizade, sempre me apoiando em todos os desafios.

À Nai por toda amizade, incentivo e carinho, por sempre acreditar em mim e apoiar todos os meus sonhos, tornando a jornada leve, divertida e feliz!

Ao Plínio por todo apoio, paciência, carinho, discussões científicas e incentivo que foram essenciais para o desenvolvimento desse projeto.

A todos os meus colegas de laboratório, em especial ao Renan, Beto, Mayara, Lêh, Ellen, Anna, Damaris, Augusto e Bárbara, por compartilharem a rotina, os desafios, e terem tornado tudo mais leve e divertido. A equipe do iGEM CNPEM Brazil 2021, pela amizade e companheirismo!

Agradeço também aos pesquisadores do LNBR que me apoiaram, tiraram dúvidas e me incentivaram ao longo do doutorado, em especial a Mari, Juh, Nádia e Cíntia. A todos os técnicos e especialistas do LNBR, em especial ao Emerson, Rebeca, Mariane, Fer, Evandro, Claudinha, Aline, Luh e Marcinha pelo companheirismo, apoio técnico e prestatividade. Agradeço também ao Ricardo, Maxuel e a Pri Tihihara por todo carinho e incentivo.

A todos os técnicos e especialistas LNBio, em especial a Gih e Celisa no auxílio dos ensaios de cristalização no ROBOLAB, e ao Felipe por todo auxílio na análise de dados de RT-qPCR.

Ao Dr. Celso Benedetti e a Jaque por todo auxílio nos experimentos com *Xanthomonas citri*, nos ensaios de nocaute gênico e nos ensaios in planta.

Ao Renan e Tati por todas as análises de espectrometria de massas, apoio e incentivo.

As equipes das linhas de luz MX2 e SAXS2 pelo suporte técnico.

Ao CNPEM pelo fornecimento de uma excelente infraestrutura para a pesquisa.

Ao Programa de Pós-Graduação em Biologia Funcional e Molecular da UNICAMP.

A toda minha família, em especial meus pais Beto e Ângela, e minha irmã, Nathy, por todo amor, ensinamentos, incentivo e pelo exemplo de força, bondade e superação. Pelo meu avô, João, que torceu por cada etapa desse processo.

Às minhas queridas amigas de república, Nat e Carol, por todos os bons momentos compartilhados.

A todos os meus amigos da Chácara Primavera e pela Rê pelo carinho, acolhimento, orações e torcida!

Resumo

Na natureza, microrganismos de diferentes nichos ecológicos desenvolveram estratégias diversas e elaboradas para a degradação de polissacarídeos complexos. Esses sistemas têm sido de grande interesse tanto por sua importância biológica, por exemplo, na reciclagem de carbono ou em interações patógeno-hospedeiro, quanto por suas aplicações biotecnológicas, como na produção avançada de biocombustíveis, nutrição animal e saúde. Bactérias fitopatogênicas do gênero *Xanthomonas*, que podem impactar significativamente a produtividade de diversas culturas agrícolas, apresentam um extenso repertório de enzimas ativas em carboidratos (CAZymes). No entanto, as estratégias moleculares utilizadas por essas bactérias para despolimerizar e captar polissacarídeos vegetais, como também as principais alterações fisiológicas desencadeadas por esses carboidratos durante o processo de infecção, ainda são parcialmente compreendidas. Sendo assim, esta tese teve como objetivo desvendar os diferentes sistemas de utilização de polissacarídeos da bactéria *X. citri* pv. *citri* cepa 306, agente causal do cancro cítrico, e revelar a resposta de expressão gênica global desta bactéria na presença de carboidratos da parede celular vegetal ou de organelas de armazenamento. Nossos resultados de RNA-seq combinados com a mineração de genoma, análises de enriquecimento e revisão de literatura revelaram três loci de utilização de polissacarídeos (PUL, do inglês *Polysaccharide Utilization Loci*) para utilização de amido, xiloglucana e galactomanana, além de indicar um possível papel de enzimas das famílias GH115 e GH2 na despolimerização de xilanas, algo ainda não reportado na literatura para esse gênero bacteriano. Nossos resultados também evidenciam que *X. citri* pode distinguir as ligações glicosídicas e decorações de cada carboidrato, como revelado pelas diferenças nos perfis de expressão gênica na presença de celobiose e amido ou xilana e arabinoxilana. Também foi observado que cada carboidrato vegetal pode atuar como um marcador espacial ou temporal, modulando as estratégias de infecção de *X. citri*. No sistema de utilização de xiloglucana, por exemplo, os produtos da degradação desse polissacarídeo foram capazes de ativar importantes fatores de virulência: o sistema de secreção do tipo III (T3SS) e os efetores do tipo III (T3E). Além disso, este trabalho também apresenta a caracterização estrutural e bioquímica da endo- β -1,4-mananase XagD, fornecendo dados moleculares sobre seu suposto papel na remodelagem de biofilme em *X. citri* e da composição ainda desconhecida do exopolissacarídeo Xag. Portanto, usando uma combinação de diferentes abordagens experimentais e computacionais, este trabalho amplia o conhecimento acerca de como carboidratos complexos podem ser despolimerizados, assimilados e utilizados como sinalizadores na modulação do metabolismo de fitopatógenos, destacando a importância dos sistemas de utilização de carboidratos complexos, em particular das CAZymes, durante o processo de infecção.

Abstract

In nature, microorganisms from different ecological niches established elaborated and diversified strategies for the degradation of complex polysaccharides. These systems have been of great interest both by their biological importance, for example, in carbon recycling or host-pathogen interactions, as by their biotechnological applications, such as in advanced biofuel production, animal nutrition, and healthcare. Phytopathogenic bacteria from the *Xanthomonas* genus, which can significantly impact the productivity of several crops, show an extensive repertoire of carbohydrate-active enzymes (CAZymes). However, the molecular strategies used by these bacteria to depolymerize and uptake plant polysaccharides, as the main physiological changes triggered by these carbohydrates during the process of infection, are still partially understood. Therefore, this thesis aimed to unravel the different polysaccharides utilization systems of the bacteria *X. citri* pv. *citri* strain 306, the causal agent of citrus canker, and shed light on the global expression response of this bacteria in the presence of carbohydrates from the plant cell wall or storage organelles. Our RNA-seq results combined with genome mining, enrichment analysis and literature curation revealed three Polysaccharide Utilization Loci (PUL) for starch, xyloglucan and galactomannan utilization, besides indicating a putative role of enzymes from families GH115 and GH2 in xylan depolymerization, which has not yet been reported in the literature for this bacterial genus. Our results also evidence that *X. citri* can distinguish the glycosidic bonds and decorations of each carbohydrate, as revealed by the differences in the genic expression profiles in the presence of cellobiose and starch or xylan and arabinoxylan. It was also observed that each plant carbohydrate can act as a spatial or temporal marker, modulating the infection strategies of *X. citri*. In the xyloglucan utilization system, for example, the products of degradation of this polysaccharide were able to activate key virulence factors: the type III secretion system (T3SS) and type III effectors (T3E). Moreover, this work also brings the structural and biochemical characterization of the endo- β -1,4-mannanase XagD, providing insights into its putative role in biofilm remodeling in *X. citri* and clues for the composition of the unknown exopolysaccharide Xag, besides revealing new structural features of the underexplored GH113 family. In conclusion, using the combination of different computational and experimental approaches, this work expands the current knowledge of how complex carbohydrates can be depolymerized, assimilated and used as signals in the modulation of phytopathogen metabolism, highlighting the importance of complex carbohydrate utilization systems, in particular CAZymes, during the infection process.

Summary

Thesis outline	13
Chapter 1.....	14
Introduction	14
Plant cell wall: a complex reservoir of carbohydrates and phenolic compounds.....	14
Polysaccharide Utilization Loci (PUL) in bacteria	16
<i>Xanthomonas</i> phytopathogens as a model of study.....	18
Host colonization by <i>Xanthomonas</i> spp. and the role of CAZymes during infection	19
Biological and biotechnological relevance of polysaccharide utilization systems.....	21
References	22
Chapter 2.....	27
Abstract	28
Introduction	29
Results	30
Plant carbohydrates trigger specific responses in <i>X. citri</i>	30
Deciphering the starch utilization machinery in <i>X. citri</i>	34
Starch inhibits sulfur metabolism and stimulates a sessile state in <i>X. citri</i>	36
<i>X. citri</i> displays an unconventional system for galactomannan utilization and a modest transcriptional response to GMOs.....	38
Xylan utilization in <i>X. citri</i> likely involves unprecedented GH115 and GH2 enzymes.....	42
Differences in heteroxylan composition are sensed by <i>X. citri</i> triggering specific responses with minimal superposition	45
Cellobiose induces the expression of GH5 endoglucanases and several TBDTs for β -glucans utilization in <i>X. citri</i>	48
Cellobiose inhibits transcriptional regulators expression and induce motility and sulfur metabolism	51
Discussion	53
Methods.....	54
Growth curves analysis	54
Hydrolysates preparation.....	54
Cloning and protein heterologous expression	55
Protein purification.....	55
RNA extraction and cDNA library preparation	56
RNA sequencing and analysis.....	56
Plate assay for carbohydrate degradation	57
References	58
Supplementary Information.....	64
Chapter 3.....	82

Introduction	84
Results	86
The XyGUL gene architecture in <i>Xanthomonas</i> diverges from Bacteroidetes	86
XyGUL endo-enzyme exploits arginine-carbohydrate interactions	88
The uptake of XyGOs is mediated by TonB-dependent transporters.....	89
XyGOs deacetylation by <i>Xanthomonas</i> involves a distinctive acetylsterase	90
<i>Xanthomonas</i> enzymatic cascade for XyGOs breakdown	93
XyG depolymerization products activate the expression of virulence factors	97
Discussion	101
Methods.....	104
Phylogenetic analysis	104
Molecular cloning and site-directed mutagenesis	105
Protein expression and purification.....	105
Analytical hydrodynamic analysis	105
Protein crystallization, X-ray data collection and structure determination	106
Glycoside hydrolase assays.....	106
<i>Arabidopsis thaliana</i> cultivation and XyGOs enzymatic extraction.....	107
Xyloglucan acetylsterase assays.....	108
<i>Xanthomonas</i> cultivations.....	109
RNA sequencing and analysis.....	110
RT-qPCR analysis.....	110
Gene knockout in <i>X. citri</i>	111
Virulence assays.....	111
Data availability	112
References	114
Reprint Permission	120
Supplementary Information.....	121
Chapter 4.....	193
Abstract	194
Introduction	194
Results and Discussion.....	196
XagD shows activity in polymeric heteromannans	196
XagD overall structure	200
Iodine ions induced the closure of a lid over the substrate-binding subsites +1 and -1 in XagD ..	202
Modifications in specific residues decrease affinity to mannose moieties at positive and negative subsites in XagD	205
Remodeling of substrate-binding cleft in XagD correlates with a possible role in EPS cleavage	208
Conclusion.....	208

Methods.....	209
Cloning, protein expression and purification	209
Substrate screening.....	209
Biochemical assays	210
Enzymatic activity evaluated by mass spectrometry.....	210
Crystallization assays, X-ray data collection and data processing	211
Small-angle X-ray scattering.....	212
References	212
Supplementary Information.....	215
Chapter 5.....	232
Concluding remarks	232
Anexo I. Termo de autorização do comitê de biossegurança	234

Thesis outline

This work has been divided into five chapters. The **Chapter 1** contains a general introduction about the complex structure and composition of the plant cell wall, the basic structure of bacterial systems for polysaccharide utilization reported so far, an overview of the key pathogenicity factors of *Xanthomonas* spp., and the biological and biotechnological relevance of CAZymes. **Chapter 2** presents a manuscript (*in preparation*) compiling the results of a genome-wide transcriptomic analysis showing the global gene expression profile of *X. citri* in response to several plant carbohydrates, unraveling the systems dedicated to the degradation and uptake of β -glucans, heteroxylans, galactomannan and starch, as well the role of their degradation products on modulating several aspects of the bacterial physiology, such as the transition between sessile and motile states. The **Chapter 3** displays a manuscript¹ in which this work contributed on revealing how the products of xyloglucan hydrolysis can activate the expression of a key virulence factor of *Xanthomonas* spp.: the type III secretion system (T3SS). The **Chapter 4** presents a manuscript (*in preparation*) about the biochemical characterization and main structural features of the β -mannanase XagD, which belongs to a cluster of biosynthesis of exopolysaccharide (EPS), providing clues for the unknown structure of the Xag-EPS and expanding the knowledge of the untapped GH113 family. The concluding remarks of this work are available in the **Chapter 5**.

Chapter 1

Introduction

Plant cell wall: a complex reservoir of carbohydrates and phenolic compounds

In nature, carbohydrates are key molecules that can display many biological functions, acting as energy storage, signaling or structural molecules. Among the most abundant sources of carbohydrates, the plant lignocellulosic biomass constitutes a carbon reservoir with great potential for the production of biofuels and other bioproducts of high commercial value that can be integrated into a circular economy^{2,3}. However, the recalcitrance of the plant cell wall acts as one of the main barriers to the production of such components, reinforcing the importance of a deep understanding of its structure and different strategies of nature to its depolymerization.

The structure of the plant cell wall consists of an intricate network of polysaccharides, containing cellulose, hemicelluloses and pectin, which may vary with the type of tissue or species⁴. These carbohydrates have a complex and heterogeneous structure and are closely linked to aromatic compounds, such as lignin and hydroxycinnamic acids (like ferulic and *p*-coumaric acid), which contribute considerably to the recalcitrance of biomass to chemical or enzymatic degradation⁵. The different compositions and proportions of its components, as well as their interactions, govern the physicochemical variations and structural characteristics of the plant cell wall⁶ and contribute to other physiological functions such as growth, defense against pathogens, mechanical resistance and intercellular communication or with the environment^{7,8}.

The hemicelluloses consist of linear or branched polysaccharides composed of different monosaccharides (e.g. D-xylose, D-galactose, D-mannose, D-glucose and L-arabinose) connected by α - or β - linkages that may contain acetylations⁹ (Figure 1). Its non-crystalline structure is interconnected with cellulose microfibrils, in addition to interacting with the pectic matrix consisting of homogalacturonans (HGA), xylogalacturonans (XGA), rhamnogalacturonans I (RGI) and rhamnogalacturonans II (RGII)¹⁰ (Figure 1). The gel phase composed by pectin is mainly present in the primary cell wall and can also interact with hemicelluloses through covalent and non-covalent interactions¹¹. Pectic polysaccharides are also present in the middle lamella (ML), which maintain cells adhered to each other, contributing to the integrity of plant tissues and organs¹² (Figure 2A).

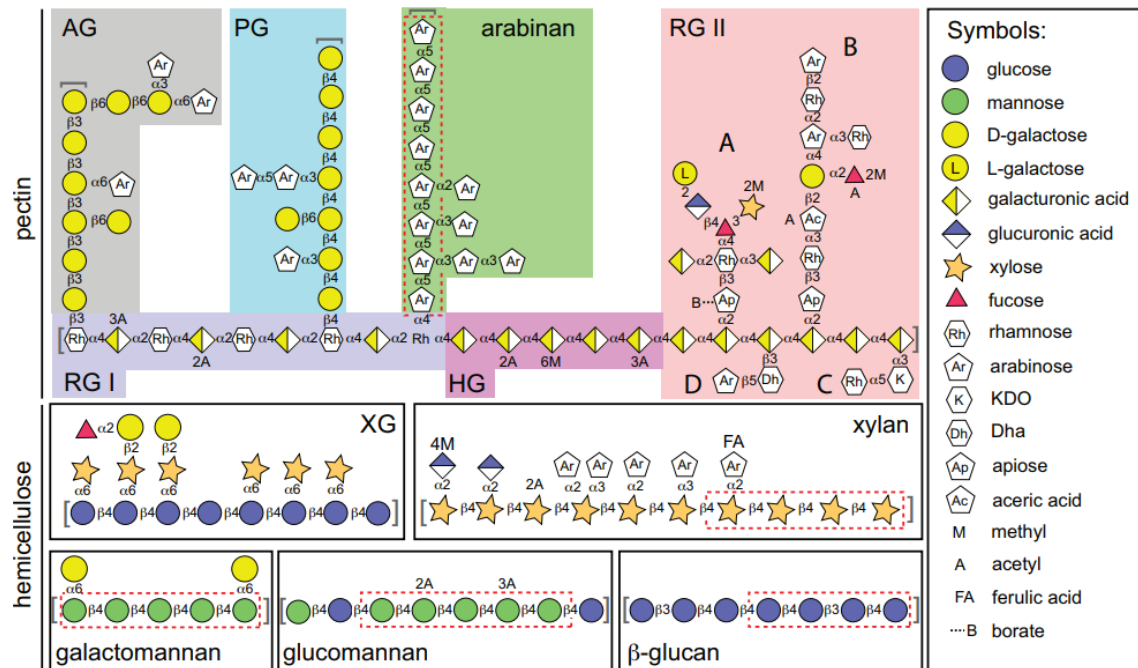


Figure 1. Structures representation of plant cell wall pectin's and hemicelluloses. The structures represent the predominant linkages and monosaccharides of each glycan and do not represent the exact structure of each carbohydrate, that may vary according to species or tissue of origin. (Figure extracted from Martens et al., 2011)¹³. AG, arabinogalactan; PG, pectic galactan; HG, homogalacturonan; RG I, rhamnogalacturonan I; RG II, rhamnogalacturonan II; XG, xyloglucan.

The main types of hemicelluloses include xyloglucans, xylans, mannans and mixed β -(1 \rightarrow 3, 1 \rightarrow 4) linkage glucans⁹. These polysaccharides also contribute to cell wall recalcitrance and require a variety of enzyme systems for their degradation¹⁴. In the primary cell wall of dicots, xyloglucans and glucuronoarabinoxylan (GAX) are mainly found^{4,15}. The primary walls are usually non-lignified and are formed during cells early development and, for some cell types (e.g. parenchyma cells) is the only wall observed⁴ (Figure 2B). In the secondary cell wall, on the other hand, the polysaccharide matrix is more closed in relation to the primary cell wall, containing few pectic polysaccharides and presenting xylans and mannans as main hemicelluloses, in addition to being lignified^{4,15} (Figure 2C). In the case of secondary walls, they are deposited in specific tissues such as xylem and sclerenchyma, providing mechanical strength and allowing the upright growth of tissues and an efficient water transport¹⁶.

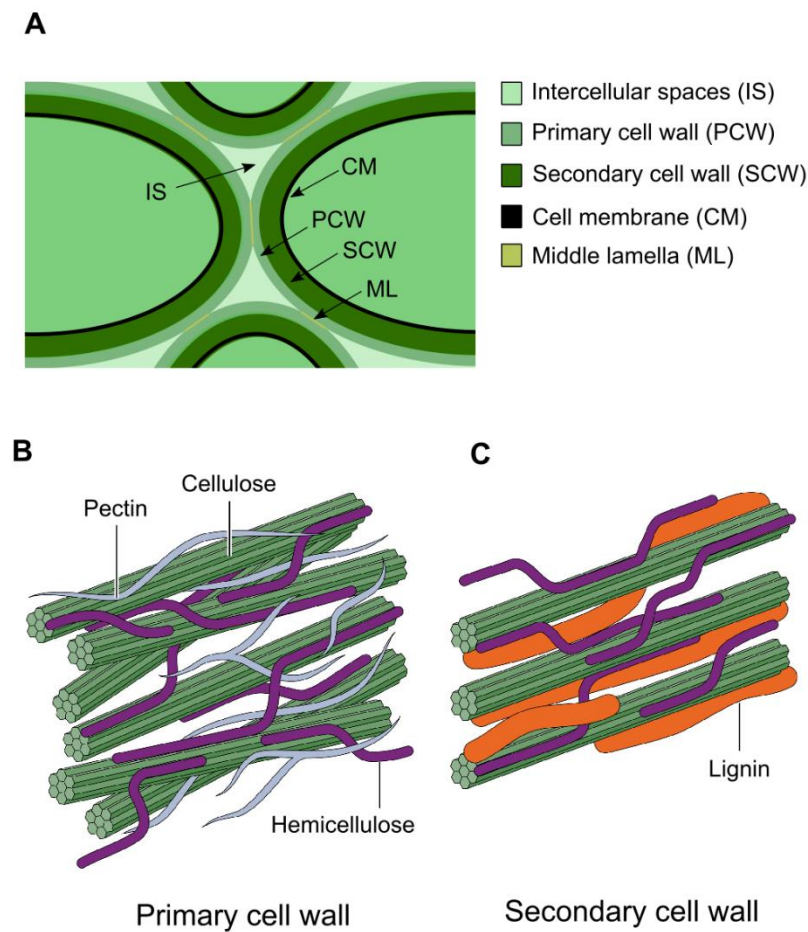


Figure 2. Schematic representation of plant cells and the plant cell wall. **A)** Simplified representation of a plant tissue showing the plant cells glued by the middle lamella and surrounded by intercellular spaces, with the deposition of the secondary cell wall between the primary cell wall and the cell membrane. **B)** Representation of the primary cell wall, showing cellulose microfibrils, hemicelluloses and pectin. **C)** Representation of the secondary cell wall, which also shows cellulose and hemicelluloses, but with a small amount of pectin and a higher content of lignin. (Figure adapted from Rytioja et al., 2014)¹⁷.

Polysaccharide Utilization Loci (PUL) in bacteria

To cope with the complex structure of different polysaccharides, several bacteria developed elaborated systems to depolymerize these substrates. Polysaccharide utilization loci (PUL) comprehends a group of co-regulated genes responsible for the coordinate hydrolysis and uptake of a complex polysaccharide^{18,19}. Its basic structure consists of genes encoding carbohydrate active enzymes (CAZymes) associated with TonB dependent transporters (TBDT), cell surface glycan-binding proteins (SGBP) and transcriptional regulators that act in

carbohydrate sensing²⁰, being commonly organized as clusters in the genome, acting or not as an operon¹⁹ (Figure 3A). The glycoside hydrolases (GHs) encoded by these loci usually display endo- or exo-activities, *i.e* cleave linkages within the polysaccharide's chains or at their extremities, and are localized in the extracellular medium, releasing oligosaccharides that are internalized by TBDTs. Next, in the periplasm, other GHs or accessory enzymes (like esterases, pectate liases, etc.) can cleave oligosaccharides and remove remaining decorations that are internalized by inner membrane transporters, such as major facilitator superfamily (MFS) transporters, and metabolized in the cytoplasm (Figure 3B).

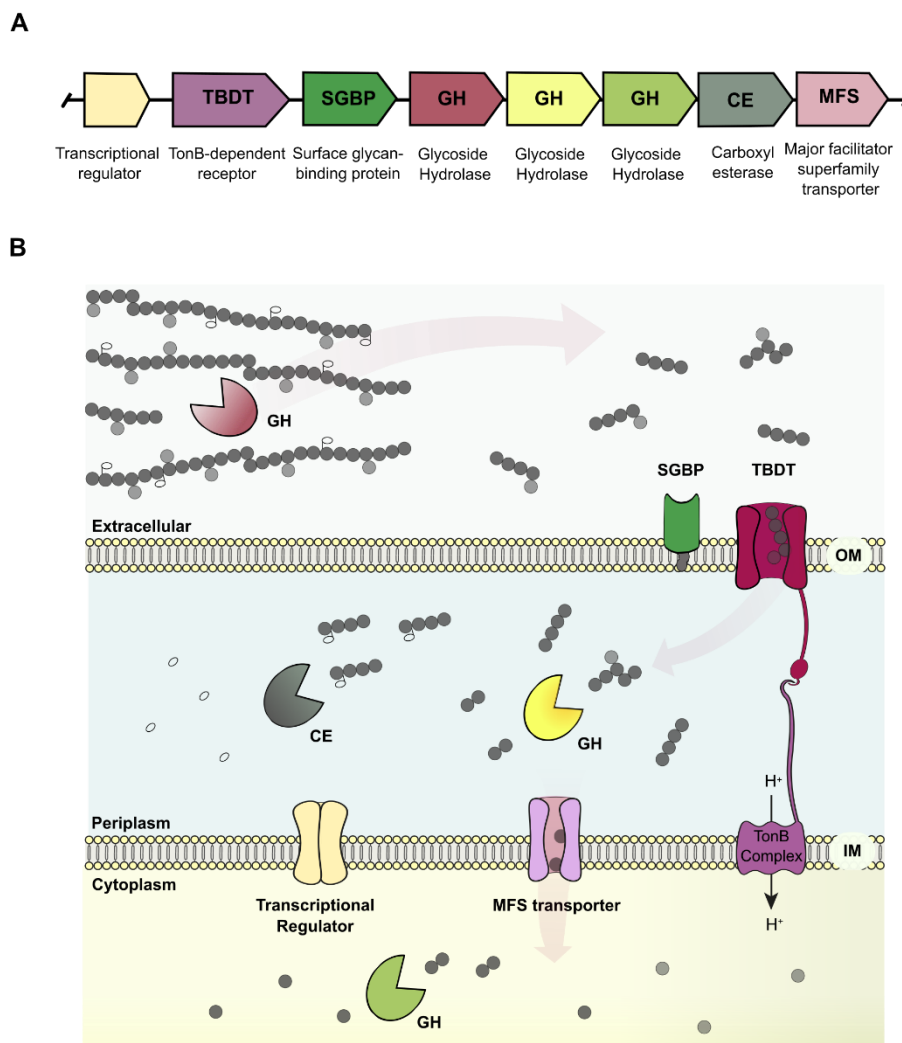


Figure 3. Schematic representation of Polysaccharide Utilization Loci (PUL) in Gram-negative bacteria. **A)** Basic structure and genetic organization of a PUL. The number and type of transporters, glycoside hydrolases and accessory enzymes are only representative and may vary according to the target polysaccharide or the bacteria genus/species. **B)** Representation of the coordinate hydrolysis and uptake of a polysaccharide by proteins encoded in a PUL. Initially, secreted endo- or exo-enzymes

recognize and cleave a polysaccharide (monosaccharides and acetate residues are represented by grey circles and empty ellipses, respectively), generating oligosaccharides that are recognized by a cell surface glycan-binding protein (SGBP) and internalized by TonB-dependent transporters (TBDT). In the periplasm, the remaining oligosaccharides are recognized by other enzymes that remove decorations and generate shorter oligosaccharides or monosaccharides that will be transported through an inner membrane transporter, like a major facilitator superfamily (MFS) transporter, to the cytoplasm. The products generated in this coordinated hydrolysis can be used as a source of carbon and energy or act as signaling molecules. Different colors of glycoside hydrolases indicate different enzymatic activities. OM: outer membrane; IM: inner membrane.

The length, composition and configuration of PULs usually varies between different microorganisms. For example, the xyloglucan PUL (XyGUL) of *Cellvibrio japonicus* lacks an extracellular endoxyloglucanase, a SGBP protein and a sensor/transcriptional regulator commonly found in other XyGULs²¹. These shorter PULs are also observed in other bacteria and can display an important role for polysaccharide degradation in specific ecological niches²⁰.

Previous studies with two human gut Bacteroides, *B. thetaiotaomicron* and *B. ovatus*, revealed that the expression of a PUL was induced only by the oligosaccharides derivate of a single polysaccharide, suggesting that a PUL is specific for the degradation of a unique carbohydrate polymer in Bacteroidetes phylum¹³. However, the mechanisms by which bacteria can recognize carbohydrates and activate the molecular machinery for their degradation and metabolism are still poorly understood. So far, PULs regulation are mainly reported to occur through SusR sensor/regulators, extracytoplasmic function sigma (ECF) factors, or hybrid two-component systems (HTCSs), although LacI, CRP, AraC (non-HTCS), SARP-OmpR, and classic TCSs regulators were also reported to play a role in PULs regulation^{20,22}. Notably, most studies related to the mechanisms of regulation and signaling of PULs were carried out for the phylum Bacteroidetes, with little information for the phylum Proteobacteria²⁰.

***Xanthomonas* phytopathogens as a model of study**

Xanthomonas comprises a genus of yellow pigmented and gram-negative bacteria belonging to the Proteobacteria phylum's gamma subdivision²³. These plant-associated bacteria are mainly pathogenic and can affect more than 400 different plant hosts, including monocots and dicots²⁴, besides being widely used for the production of xanthan gum on an industrial scale²⁵. Although they can affect a wide variety of crops, each species show high specificity for a limited number of hosts and can generate infections in the vascular regions of the xylem or

occupy the intercellular spaces of the parenchyma mesophyll²⁶. In general, the diseases caused by *Xanthomonas* spp. have a significant impact in the productivity of several important crops like rice²⁷, cabbage²⁸, banana²⁹, orange²³ and others, affecting the food security in several countries. In this way, due to a high number of hosts and the different types of tissues these bacteria can infect, the genus *Xanthomonas* is an interesting model to be explored that can largely contribute to a deeper understanding of host-pathogen interactions and reveal new molecular strategies for polysaccharides depolymerization that could be exploited for biotechnological applications.

Notably, *Xanthomonas* spp. usually shows an enrichment of TBDT in their genome, with many of them been associated to CAZymes and transcriptional regulators, indicating a capacity to degrade a variety of plant polysaccharides³⁰. Besides showing a high potential for polysaccharides degradation, only the carbohydrate utilization systems for xylan^{31,32} and N-glycans^{33,34} have been characterized in the genus *Xanthomonas* so far. These systems, also denominated previously as CUT (Carbohydrate Utilization locus-containing TBDT)³¹, are similar to Bacteroidetes systems but commonly lack cell surface glycan-binding proteins (SGBP). In the genus *Xanthomonas* there is some studies related to carbohydrate metabolism like glucose³⁵ and galactose³⁶, but few information about complex carbohydrates sensing, hydrolysis and uptake. Notably, the carbon metabolism is commonly associated to virulence mechanisms in these bacteria, influencing in the exopolysaccharides (EPS) biosynthesis in biofilm formation, for example³⁷. Many plant cell wall degrading enzymes (PCWEs) also show a notorious role in *Xanthomonas* infection development at the host (see next section), reinforcing the importance of a deep understanding of PCWEs biochemical and biological functions in the context of different PULs.

Host colonization by *Xanthomonas* spp. and the role of CAZymes during infection

In nature, *Xanthomonas* spp. dissemination can occur mainly via contaminated seeds, through the wind, rain or water splash²⁶. In the initial phases of the infection, the bacteria grow in an epiphytic mode on the surface of leaves and fruits and can attach to the surface through adhesins proteins, polysaccharides and the type IV pilus^{26,38}. After a few days, the bacteria can migrate to the entry points in the host, forming microcolonies around the wounds, hydathodes or stomata, allowing the entering of these bacteria in the internal tissues to further start the endophytic stage of infection^{39,40}.

Most of *Xanthomonas* spp. can display two types of infection: vascular or mesophyllic⁴¹. In the first type, bacteria can enter through hydathodes or wounds to colonize the xylem vessels, later occluding the vascular systems and generating necrosis, as observed for *X. campestris* pv. *campestris* causing black rot in brassicas^{28,41}. In the second type, the bacteria can access the intercellular spaces of the mesophyll parenchyma through stomata or wounds, as observed for *X. citri* pv. *citri*, the causal agent of citrus canker^{23,38}. Notably, in cassava bacterial blight, *Xanthomonas manihotis* can enter the leaves through hydathodes and stomata, infecting both the intercellular spaces and the vascular tissues⁴².

During the different stages of infection, *Xanthomonas* spp. can display many mechanisms to colonize the host, uptake nutrients and overcome the host's immune response. Among the several virulence factors used to succeed during the infection process, *Xanthomonas* spp. can synthesize exopolysaccharides (e.g., xanthan gum) for the formation of a biofilm matrix, contributing for cell-cell attachment or adhesion to surfaces and creating a protection barrier against antimicrobials compounds or host responses^{43,44}. Furthermore, the type IV pilus (T4P), besides allowing the twitching motility, were also reported to contribute with the formation of mushroom-like structures in *X. citri* pv. *citri* mature biofilms and with the adherence to plant tissues⁴⁵. The biofilm formation/dispersal is mainly coordinated by quorum sensing and the diffusible signal factor (DSF), with the movements of the bacteria across the environment or in the host tissues directed by chemotaxis in coordination with flagella-dependent or independent movements^{46,47}.

The secretions systems of *Xanthomonas* spp. are also important virulence factors that play a notorious role in host colonization. For example, the type I secretion system (T1SS) can delivery bacteriocins, proteins containing a nonapeptide repeat in toxins (RTX) and adhesins, allowing the adhesion to the host cells⁴⁸. In a similar way, the type V secretion system (T5SS) also display a role in adhesion to host and non-host cells through non-fimbrial adhesins, besides contributing to cell-cell attachment in biofilm formation⁴⁸. Regarding the interaction with the microflora phyllosphere, *Xanthomonas* spp. can use the type IV and type VI secretion systems (T4SS and T6SS) in the competition with other microorganisms or eukaryotic predators⁴⁸. The T4SS is responsible for delivering toxins to other bacteria, acting as a nano weapon to kill bacterial competitors, besides showing a possible role in the mobilization of genetic material for bacterial conjugation^{48,49}. The T6SS is responsible for providing resistance for the soil amoeba *Dictyostelium discoideum*, and was reported to be induced during the epiphytic colonization of *X. citri*^{50,51}.

The type III secretion system (T3SS) is a key virulence factor encoded by the *hrp* cluster in *Xanthomonas* spp.⁵². The T3SS composes a pilus that can cross the plant cell wall (PCW) and translocate several effectors into the host cell, modulating its immune response and metabolism to allow the infection development^{53,54}. The disruption of the PCW for the T3SS-pilus assembly, however, is a challenging process since the PCW shows at least 100 nm-wide and potentially requires the action of CAZymes to assist the hydrolysis of plant polysaccharide matrix⁵⁵. Notably, other CAZymes like the putative lytic transglycosylase Hpa2 from *X. oryzae*, classified in family GH23⁵⁶, can assist the T3SS-pilus assembly in the bacterial plasma membrane^{57,58}.

During the process of infection, the role of CAZymes in *Xanthomonas* spp. can go beyond the plant cell wall degradation of the host. For example, the enzyme ManA is considered the major mannanase in *Xanthomonas* spp. and was previously reported to show a biological role in biofilm dispersal in *X. campestris*⁵⁹. A recent report also evidenced the role of a single cellobiohydrolase, CbsA on determining the tissue specificity of *Xanthomonas* spp., acting as a switch from non-vascular to vascular pathogenesis^{60,61}. In many phytopathogens as well in *Xanthomonas* spp., CAZymes are mainly secreted through the type II secretion system (T2SS) or through outer membrane vesicles (OMVs)⁶². The decreased virulence of mutants lacking T2SS suggests a reduction in CAZymes secretion or of unknown virulence effectors, and indicates the relevance of this system for a successful infection^{62,63}.

Biological and biotechnological relevance of polysaccharide utilization systems

In bacteria, carbohydrates were already reported to act as signaling molecules that allow the adaptation according to the availability of nutrient sources or the control of different steps of the infection⁶⁴. For example, the phytopathogen *Agrobacterium tumefaciens* can recognize various host sugars using a periplasmic-binding protein encoded by the chromosomal virulence gene *E* (*chvE*), indicating that this bacterium probably uses the host carbohydrates as spatial and temporal markers among the different tissues to regulate the steps of infection⁶⁵. Similarly, the human pathogen *Streptococcus pneumoniae* also shows genes involved in carbohydrate uptake and metabolism connected to virulence mechanisms, like the use of galactose as a marker and an essential nutrient for the development of infection in the respiratory tract⁶⁶. Similarly, previous reports in *X. campestris* pv. *vesicatoria* revealed the activation of virulence factor T3SS in the presence of simple sugars such as saccharose and fructose⁶⁷. However, despite the notorious influence of carbohydrates in the process of infection of several bacteria,

information about how plant polysaccharides can modulate physiological processes in *Xanthomonas* spp. remains so far elusive.

Besides the importance of understanding the influence of complex carbohydrates in host-pathogen interactions, the knowledge of carbohydrate degradation systems can also promote biotechnological advances. For example, the understanding of these systems can be applied in the generation of functional oligosaccharides important for human health⁶⁸ or in the tailoring of enzymatic cocktails for lignocellulose biomass hydrolysis⁶⁴. Furthermore, the deep knowledge of all the protein required to hydrolyze and uptake different carbohydrates can also be incorporated into the rational design of strains or synthetic biochemical pathways for bioproducts generation using complex carbohydrates as a feedstock⁶⁹.

Thus, considering the biological and biotechnological relevance of understanding the mechanisms of sensing, hydrolysis and uptake of polysaccharides, this study used the causal agent of citrus canker, *X. citri* pv. *citri* strain 306, as a study model to evaluate the transcriptional responses triggered in this bacterium by complex carbohydrates derived from the plant cell wall or reserve organelles, expanding the current knowledge about the carbohydrate degradation systems of this bacterium and the main influence of their degradation products in modulating bacterial physiological processes.

References

1. Vieira, P. S. *et al.* Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* **12**, 1–15 (2021).
2. Isikgor, F. H. & Becer, C. R. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym. Chem.* **6**, 4497–4559 (2015).
3. Lee, S. Y. *et al.* A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2**, 18–33 (2019).
4. Harris, P. J. & Stone, B. A. Chemistry and Molecular Organization of Plant Cell Walls. *Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy* (2009). doi:10.1002/9781444305418.ch4.
5. de Oliveira, D. M. *et al.* Ferulic acid: A key component in grass lignocellulose recalcitrance to hydrolysis. *Plant Biotechnol. J.* **13**, 1224–1232 (2015).
6. Pattathil, S., Hahn, M. G., Dale, B. E. & Chundawat, S. P. S. Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEXTM-pretreated biomass. *J. Exp. Bot.* **66**, 4279–4294 (2015).
7. Minic, Z. & Jouanin, L. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* **44**, 435–449 (2006).

8. Harris, P. J. & Smith, B. G. Plant cell walls and cell-wall polysaccharides: Structures, properties and uses in food products. *Int. J. Food Sci. Technol.* **41**, 129–143 (2006).
9. Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289 (2010).
10. Keegstra, K. Plant cell walls. *Plant Physiol.* **154**, 483–486 (2010).
11. Jarvis, M. C. Structure and properties of pectin gels in plant cell walls. *Plant. Cell Environ.* **7**, 153–164 (1984).
12. Zamil, M. S. & Geitmann, A. The middle lamella - More than a glue. *Phys. Biol.* **14**, (2017).
13. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, (2011).
14. Gilbert, H. J. The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.* **153**, 444–455 (2010).
15. Loix, C. *et al.* Reciprocal interactions between cadmium-induced cell wall responses and oxidative stress in plants. *Front. Plant Sci.* **8**, 1–19 (2017).
16. Zhong, R. & Ye, Z. H. Secondary cell walls: Biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* **56**, 195–214 (2015).
17. Rytioja, J. *et al.* Plant-Polysaccharide-Degrading Enzymes from Basidiomycetes. *Microbiol. Mol. Biol. Rev.* **78**, 614–649 (2014).
18. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J. Biol. Chem.* **281**, 36269–36279 (2006).
19. Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647–655 (2015).
20. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, 1–15 (2017).
21. Larsbrink, J. *et al.* A complex gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio japonicus*. *Mol. Microbiol.* **94**, 418–433 (2014).
22. Schwalm, N. D. & Groisman, E. A. Navigating the Gut Buffet: Control of Polysaccharide Utilization in Bacteroides spp. *Trends Microbiol.* **25**, 1005–1015 (2017).
23. Brunings, A. M. & Gabriel, D. W. *Xanthomonas citri*: breaking the surface. *Mol. Plant Pathol.* **4**, 141–157 (2003).
24. Timilsina, S. *et al.* *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nat. Rev. Microbiol.* **18**, 415–427 (2020).
25. Palaniraj, A. & Jayaraman, V. Production, recovery and applications of xanthan gum by *Xanthomonas campestris*. *J. Food Eng.* **106**, 1–12 (2011).
26. An, S. Q. *et al.* Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiol. Rev.* **44**, 1–32 (2019).
27. Zhang, H. & Wang, S. Rice versus *Xanthomonas oryzae* pv. *oryzae*: a unique pathosystem. *Curr. Opin. Plant Biol.* **16**, 188–195 (2013).
28. Vicente, J. G. & Holub, E. B. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Mol. Plant Pathol.* **14**, 2–18 (2013).
29. Tripathi, L. *et al.* *Xanthomonas* Wilt: A Threat to Banana Production in East and Central Africa.

- Plant Dis.* **93**, 440–451 (2009).
30. Blanvillain, S. *et al.* Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. *PLoS One* **2**, (2007).
 31. Déjean, G. *et al.* The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv. *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytol.* **198**, 899–915 (2013).
 32. Chow, V. *et al.* Xylan utilization regulon in *Xanthomonas citri* pv. *citri* strain 306: gene expression and utilization of oligoxylosides. *Appl. Environ. Microbiol.* **81**, 2163–2172 (2015).
 33. Boulanger, A. *et al.* The plant pathogen *Xanthomonas campestris* pv. *campestris* exploits N-Acetylglucosamine during infection. *MBio* **5**, 1–13 (2014).
 34. Dupoirson, S. *et al.* The N-Glycan cluster from *Xanthomonas campestris* pv. *campestris*: a toolbox for sequential plant N-Glycan processing a toolbox for sequential plant N-Glycan processing. *J. Biol. Chem.* **290**, 6022–6036 (2015).
 35. Schatschneider, S. *et al.* Metabolic flux pattern of glucose utilization by *Xanthomonas campestris* pv. *campestris*: prevalent role of the Entner–Doudoroff pathway and minor fluxes through the pentose phosphate pathway and glycolysis. *Mol. Biosyst.* **10**, 2663–2676 (2014).
 36. Serrania, J., Vorhölter, F. J., Niehaus, K., Pühler, A. & Becker, A. Identification of *Xanthomonas campestris* pv. *campestris* galactose utilization genes from transcriptome data. *J. Biotechnol.* **135**, 309–317 (2008).
 37. Kim, S. Y., Lee, B. M. & Cho, J. Y. Relationship between glucose catabolism and xanthan production in *Xanthomonas oryzae* pv. *oryzae*. *Biotechnol. Lett.* **32**, 527–531 (2010).
 38. Jacques, M. A. *et al.* Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* **54**, 163–187 (2016).
 39. Cubero, J., Gell, I., Johnson, E. G., Redondo, A. & Graham, J. H. Unstable green fluorescent protein for study of *Xanthomonas citri* subsp. *citri* survival on citrus. *Plant Pathol.* **60**, 977–985 (2011).
 40. Cerutti, A. *et al.* Immunity at cauliflower hydathodes controls systemic infection by *Xanthomonas campestris* pv. *campestris*. *Plant Physiol.* **174**, 700–716 (2017).
 41. Ryan, R. P. *et al.* Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **9**, 344–355 (2011).
 42. López, C. E. & Bernal, A. J. Cassava Bacterial Blight: Using Genomics for the Elucidation and Management of an Old Problem. *Trop. Plant Biol.* **5**, 117–126 (2012).
 43. Rigano, L. A. *et al.* Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *Mol. Plant-Microbe Interact.* **20**, 1222–1230 (2007).
 44. Crossman, L. & Dow, J. M. Biofilm formation and dispersal in *Xanthomonas campestris*. *Microbes Infect.* **6**, 623–629 (2004).
 45. Dunger, G., Guzzo, C. R., Andrade, M. O., Jones, J. B. & Farah, C. S. *Xanthomonas citri* subsp. *citri* type IV pilus is required for twitching motility, biofilm development, and adherence. *Mol. Plant-Microbe Interact.* **27**, 1132–1147 (2014).
 46. He, Y. W. & Zhang, L. H. Quorum sensing and virulence regulation in *Xanthomonas campestris*. *FEMS Microbiol. Rev.* **32**, 842–857 (2008).
 47. Moreira, L. M. *et al.* Chemotactic signal transduction and phosphate metabolism as adaptive strategies during citrus canker induction by *Xanthomonas citri*. *Funct. Integr. Genomics* **15**, 197–

- 210 (2015).
48. Alvarez-Martinez, C. E. *et al.* Secrete or perish: The role of secretion systems in *Xanthomonas* biology. *Comput. Struct. Biotechnol. J.* **19**, 279–302 (2021).
 49. Souza, D. P. *et al.* Bacterial killing via a type IV secretion system. *Nat. Commun.* **6**, 1–9 (2015).
 50. Ceseti, L. M. *et al.* The *Xanthomonas citri* pv. *citri* Type VI Secretion System is induced during epiphytic colonization of citrus. *Curr. Microbiol.* **76**, 1105–1111 (2019).
 51. Bayer-Santos, E. *et al.* *Xanthomonas citri* T6SS mediates resistance to *Dictyostelium* predation and is regulated by an ECF σ factor and cognate Ser/Thr kinase. *Environ. Microbiol.* **20**, 1562–1575 (2018).
 52. Arlat, M., Gough, C. L., Barber, C. E., Boucher, C. & Daniels, M. J. *Xanthomonas campestris* contains a cluster of *hrp* genes related to the larger *hrp* cluster of *Pseudomonas solanacearum*. *Mol. plant-microbe Interact.* **4**, 593–601 (1991).
 53. Büttner, D. & Bonas, U. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* **34**, 107–133 (2010).
 54. Kay, S. & Bonas, U. How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* **12**, 37–43 (2009).
 55. Ji, H. & Dong, H. Key steps in type III secretion system (T3SS) towards translocon assembly with potential sensor at plant plasma membrane. *Mol. Plant Pathol.* **16**, 762–773 (2015).
 56. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
 57. Li, Y. R. *et al.* Hpa2 required by HrpF to translocate *Xanthomonas oryzae* transcriptional activator-like effectors into rice for pathogenicity. *Appl. Environ. Microbiol.* **77**, 3809–3818 (2011).
 58. Zhang, J., Wang, X., Zhang, Y., Zhang, G. & Wang, J. A conserved Hpa2 protein has lytic activity against the bacterial cell wall in phytopathogenic *Xanthomonas oryzae*. *Appl. Microbiol. Biotechnol.* **79**, 605–616 (2008).
 59. Dow, J. M. *et al.* Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10995–11000 (2003).
 60. Redkar, A., Sabale, M. & Di Pietro, A. A ‘Hydrolase Switch’ for Vascular Specialization in Plant Pathogenic Bacteria. *Trends Plant Sci.* **26**, 427–429 (2021).
 61. Gluck-Thaler, E. *et al.* Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles. *Sci. Adv.* **6**, 1–11 (2020).
 62. Solé, M. *et al.* *Xanthomonas campestris* pv. *vesicatoria* secretes proteases and xylanases via the Xps type II secretion system and outer membrane vesicles. *J. Bacteriol.* **197**, 2879–2893 (2015).
 63. Szczesny, R. *et al.* Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria*. *New Phytol.* **187**, 983–1002 (2010).
 64. Østby, H., Hansen, L. D., Horn, S. J., Eijsink, V. G. H. & Várnai, A. Enzymatic processing of lignocellulosic biomass: principles, recent advances and perspectives. *Journal of Industrial Microbiology and Biotechnology* vol. 47 (Springer International Publishing, 2020).
 65. Hu, X., Zhao, J., DeGrado, W. F. & Binns, A. N. *Agrobacterium tumefaciens* recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. *Proc. Natl. Acad. Sci.*

- U. S. A.* **110**, 678–683 (2013).
66. Paixão, L. *et al.* Host glycan sugar-specific pathways in streptococcus pneumonia: Galactose as a key sugar in colonisation and infection. *PLoS One* **10**, 1–33 (2015).
 67. Wengelnik, K., Marie, C., Russel, M. & Bonas, U. Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction. *J. Bacteriol.* **178**, 1061–1069 (1996).
 68. Qiang, X., YongLie, C. & QianBing, W. Health benefit application of functional oligosaccharides. *Carbohydr. Polym.* **77**, 435–441 (2009).
 69. Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D. & Martín, H. G. Synthetic and systems biology for microbial production of commodity chemicals. *npj Syst. Biol. Appl.* **2**, 1–11 (2018).

Chapter 2

Type of chapter: Research Article

Current status: In preparation

Transcriptional landscape of *Xanthomonas citri* in response to plant carbohydrates: depicting polysaccharide utilization systems and pathogenicity modulation

Isabela Mendes Bonfim^{1,2}, Douglas Alvarez Paixão¹, José Alberto Diogo^{1,2}, Joaquim Martins Junior¹, Gabriela Felix Persinoti¹, Priscila Oliveira de Giuseppe^{1,*} and Mário Tyago Murakami^{1,*}

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPem), Zip Code 13083-970, Campinas, Sao Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

*Correspondence to priscila.giuseppe@lnbr.cnpem.br and mario.murakami@lnbr.cnpem.br

Abstract

Some phytopathogens are outfitted with a broad and diverse repertoire of enzymatic systems that enable the breakdown and utilization of host polysaccharides as a source of carbon, energy, and stimuli. However, the functional assignment of these enzymatic systems and the influence of their products on modulating pathogen behavior during host colonization is yet poorly comprehended. In this study, we performed RNA-seq analyses to provide a comprehensive, genome-wide view of the transcriptional response of the phytopathogen *Xanthomonas citri* pv. *citri* 306 to complex polysaccharides from the plant cell wall or storage organelles. Combining transcriptional and genome mining analyses, we systematically evaluated the loci employed by *X. citri* to uptake and depolymerize complex plant polysaccharides (starch, galactomannan, xylan, arabinoxylan, and β -glucans). Co-regulation and gene ontology enrichment analyses showed contrasting responses when comparing carbohydrates derived from cellulose and hemicelluloses with starch. Cellobiose, a depolymerization product of plant cell wall β -glucans, upregulates genes related to flagellum assembly and type IV pili, inducing a motile state. In contrast, the storage polysaccharide starch suppresses genes related to chemotaxis, flagellum assembly and biofilm dispersion, favoring a sessile state. Taken together, these results unravel that besides using host carbohydrates as a source of carbon and energy, *X. citri* also exploits them as signaling molecules for sensing the environment, adapting their metabolism and controlling transitions between motile and sessile states for a successful host colonization.

Keywords: RNA-seq, transcriptome, regulation, *Xanthomonas*, phytopathogens, CAZyme, polysaccharide utilization loci, plant polysaccharides.

Introduction

Nutrients are generally seen as a source of carbon and energy for living cells. However, their roles in life go far beyond that, as they trigger responses in the cells related not only to their uptake and metabolism but also to other physiologic processes important for cell survival, adaptation, and growth. This has been clearly evidenced in pathogen-host interactions, with several studies showing that pathogens from either animals or plants exploit nutrients and other signals derived from the host to regulate their virulence and adapt their metabolism for a successful colonization^{1,2}.

Simple sugars, amino acids and organic acids are usually seen as the main carbon sources available in nutritional niches in plants². However, plant polysaccharides can also provide carbon, energy, and stimuli for phytopathogens adapted to their utilization, although this is still poorly understood. Plant polysaccharides display a variety of structures defined by their monomer types and glycosidic linkages. Some of them are homopolymers either linear (ex. cellulose) or branched (ex. starch), whereas others are heteropolymers with complex structures, such as the hemicelluloses arabinoxylan and xyloglucan. In the plant cell wall, cellulose and hemicelluloses contribute to forming a complex structural matrix with pectin and lignin to support plant growth and act as a defense barrier against microbial attack³. At plant organelles, polysaccharides such as starch and mannans⁴ function as a reserve material to feed cells and adjust osmotic potential when needed⁵.

Molecular strategies for the utilization of plant polysaccharides have been well investigated in members of the Gram-negative phylum *Bacteroidetes*⁶, especially those that inhabit the human gut, but are less understood in phytopathogenic bacteria such as those from the genus *Xanthomonas*⁷, which belongs to the phylum *Proteobacteria*. These strategies include multi-enzymatic systems for polysaccharide depolymerization, transmembrane transporters to uptake their fragments and metabolic pathways to transform their monomers into energy and building blocks for biosynthetic pathways⁷. In Gram-negative bacteria, genes related to the uptake and depolymerization of a specific polysaccharide are usually grouped in genomic regions named polysaccharide utilization loci (PUL)⁷. Computational tools such as Polysaccharide-Utilization Loci DataBase PULDB (<http://www.cazy.org/PULDB/>)⁸ and dbCAN-PUL (https://bcb.unl.edu/dbCAN_PUL/)⁹ have facilitated the identification of PULs in bacterial genomes, but the functional assignment of these predictions usually requires further experimental validation. This because, the modularity of these systems has allowed the evolution of a variety of possible combinations to enable the utilization of a given

polysaccharide, whose complexity often dictates that of the corresponding PUL⁷. Moreover, there are polysaccharides that share structural similarities with each other and polysaccharides whose molecular structures are still unknown¹⁰. Although the enzymes that cleave polysaccharides have been systematically grouped in families in the Carbohydrate Active Enzymes database (CAZy, <http://www.cazy.org/>)¹¹, several families comprise enzymes with different activities, whereas the same activity sometimes is found in different families, hampering the functional assignment of PULs in microbes more distant to those taken as a validated reference.

In this work, we use *Xanthomonas citri* pv. *citri* 306 (hereafter *X. citri*), the causal agent of citrus canker, as a model to investigate, on a genome-wide scale, the molecular adaptations that some phytopathogens display to exploit plant polysaccharides as a source of carbon, energy and signals to modulate their metabolism and behavior during colonization. In a previous study, we demonstrated the molecular mechanisms *X. citri* pv. *citri* strain 306 employs to use xyloglucan, a complex polysaccharide from the plant cell wall, as a nutrient and a signaling source to modulate its virulence traits¹², adding to other efforts to understand the polysaccharide utilization systems of the genus *Xanthomonas*^{13,14,15,16}. Here, we expand our RNA-seq studies to other plant polysaccharides, providing a genome-wide transcriptome landscape of *X. citri* when exposed to the chemical complexity of carbohydrates derived from the plant cell wall or storage organelles, unraveling the transcriptional responses of this phytopathogen to each one of these carbohydrates to obtain carbon and energy for growth and to ensure a successful colonization of the host.

Results

Plant carbohydrates trigger specific responses in *X. citri*

To investigate the transcriptional responses of *X. citri* to different plant carbohydrates and to identify the enzymatic systems used by this bacterium to break down and utilize these glycans, we performed RNA sequencing (RNA-seq) studies using minimal medium XVM2m¹⁷ containing starch, a typical storage polysaccharide, or oligosaccharides from β -glucans, xylan, arabinoxylan or galactomannan as the sole carbohydrate source (Supplementary Table 1). The expression profiles of *X. citri* in the different carbohydrate sources were compared to XVM2m medium supplemented with glucose, showing a total of 11 to 150 differentially expressed genes (DEGs) in the analyzed conditions (Supplementary Data 1). When comparing the global gene

expression profiles, we noticed that all conditions display a subgroup of carbohydrate-specific DEGs, with the cellobiose condition showing the highest amount of total and unique upregulated genes (Figure 1A) and the starch condition revealing the highest amount of total and unique downregulated genes (Figure 1B).

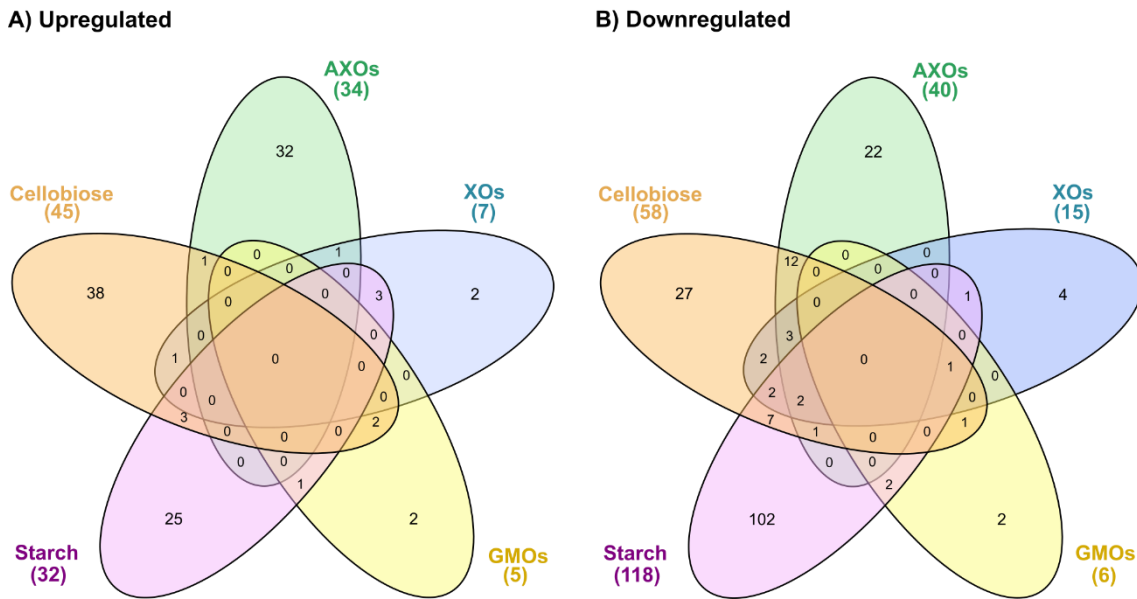


Figure 1. Venn diagrams comparing the distribution of unique and shared (A) upregulated and (B) downregulated differentially expressed genes in each condition. The differential expression in each condition was compared to XVM2m medium with glucose at the cut off $|\log_2 \text{ fold change}| \geq 1$ and $p \text{ adjusted} \leq 0.05$. The Venn diagram plots were generated using the InteractiVenn web-based tool¹⁸ (<http://www.interactivenn.net/>). AXOs: arabinoxylan oligosaccharides; XOs: xylan oligosaccharides; GMOs: galactomannan oligosaccharides.

Next, co-expression and gene set enrichment analyses were performed to better understand the main biological functions influenced by the evaluated plant carbohydrates, assuming that genes with a similar expression pattern show related biological roles¹⁹. Five modules of co-expressed genes added by a module of not correlated (NC) genes were identified (Figure 2), showing at least one significantly enriched pathway in the Over Representation Analyses (ORA), except for module 1 (M1) (Figure 3).

Overall, we observed that the starch-triggered response seems to be in the opposite direction of those triggered by the oligosaccharides derived from the plant cell wall, which correlates with their different molecular structures and spatial locations in plant cells (Figure

2). Modules M2 and M5, which include genes involved in chemotaxis and sulfur metabolism, were less active in the starch condition but more active in the cellobiose, XOs and GMOs conditions. On the other hand, the transcriptional activity of modules M4 and NC, which include pathways related to amino acids, sugar and starch metabolism, was higher in the starch condition compared to the other analyzed carbohydrates (Figures 2 and 3).

Intriguingly, the conditions XOs and AXOs, both representative of xylan oligosaccharides, displayed dissimilarities in the responses observed for most modules, which might be associated with the higher content of arabinose in AXOs (Figures 2 and 3). Together, these analyses indicate that differences either in the monomer's composition or in the type of linkages present in the substrates, for example, α -glucosyl in starch and β -glucosyl in cellobiose, are sensed by *X. citri*, modulating the intensity and the type of the transcriptional response.

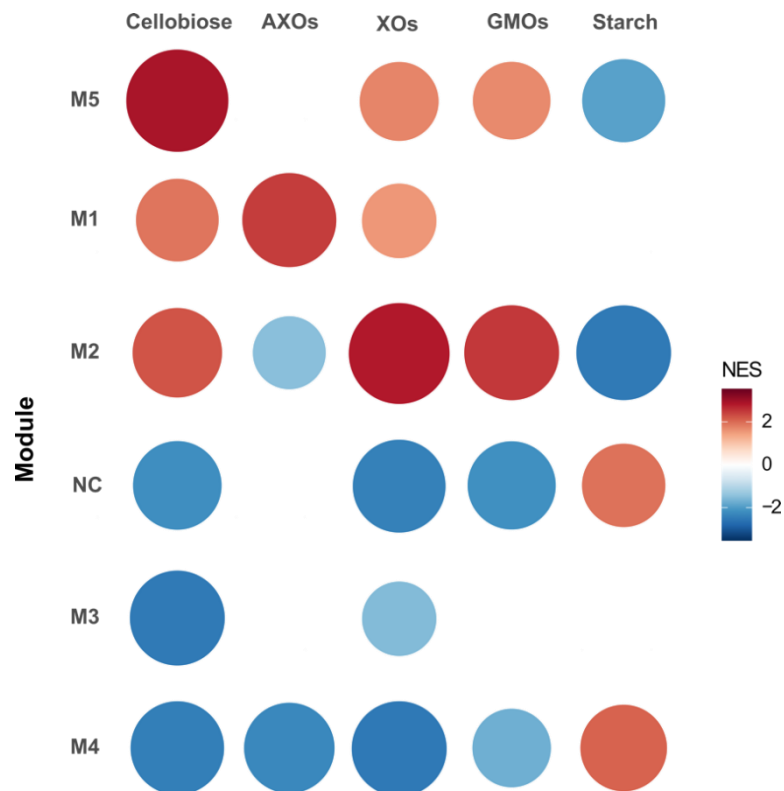


Figure 2. Gene set enrichment analysis showing module activity on each condition. In these analyses, genes from co-expression modules were treated as gene sets and the z-score normalized gene expression of the samples within each condition was treated as rankings on the analysis²⁰. The size and intensity of the circles correspond to the normalized enrichment score (NES) for the module in each condition. Positive NES reflects transcriptional activity above the median, whereas negative NES

correspond to transcriptional activity below the median in each condition. This plot was generated using the CEMiTool package²⁰.

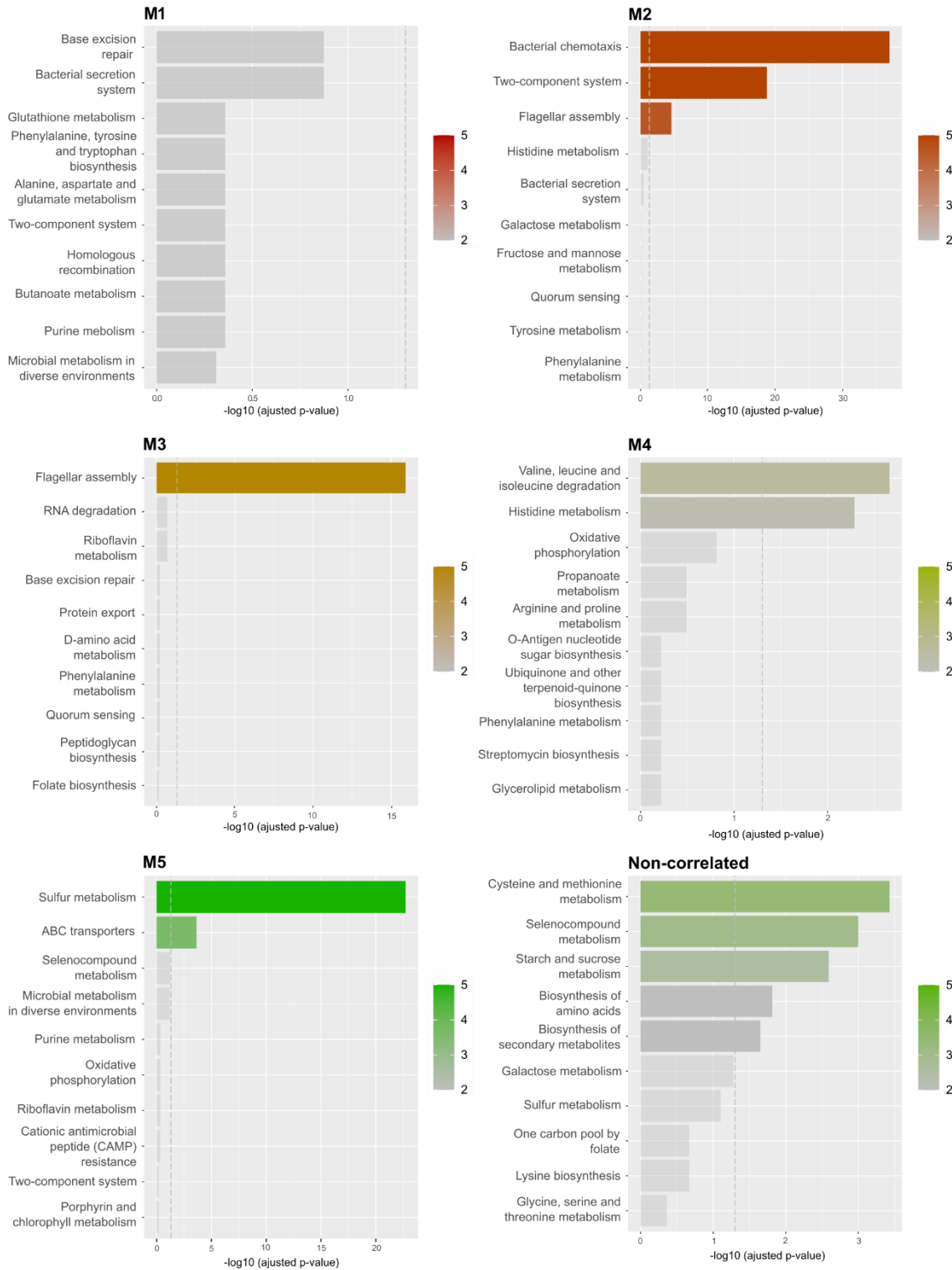


Figure 3. Over representation analysis to identify biological functions enriched in each module. The x-axis corresponds to $-\log_{10}$ of the Benjamini-Hochberg (BH)-adjusted p-value, and the dashed vertical line indicates the BH-adjusted p-value threshold of 0.05. The color gradients are proportional to the adjusted p-value. This plot was generated using the CEMiTool package²⁰.

Deciphering the starch utilization machinery in *X. citri*

Starch is a storage carbohydrate in plants and it is composed of amylose (long linear chains of α -(1 \rightarrow 4)-linked D-glucose residues (Figure 4A)) and amylopectin (short chains of α -(1 \rightarrow 4)-linked D-glucose residues with α -(1 \rightarrow 6)-branch linkages (Figure 4B))²¹. Their granules are stored in organelles named amyloplasts and can be found within cells from leaves, seeds, fruits, stems and roots²².

Searches in the *X. citri* genome for genes potentially involved with starch depolymerization revealed 1 to 3 belonging to the families GH4, GH15 and GH97, and notably 12 belonging to the family GH13 (Supplementary Table 2). From these genes, only some members of families GH13 (XAC0798, XAC2596 and XAC2602) and GH97 (XAC2599) were up-regulated in the starch condition (Figure 4C, Supplementary Data 1), supporting its role in starch depolymerization. Some of them are next to the loci encoding for a transcriptional regulator (XAC2595), an inner membrane MFS transporter (XAC2597) and a TonB-dependent transporter (TBDT, XAC2600), with the last two upregulated in presence of starch (Figures 4C and D). Thus, our results indicate that the gene cluster XAC2595-XAC2602 along with the gene XAC0798, which encodes the α -amylase Amy²³, composes the Starch utilization system (Sus) of *X. citri* (Figure 4D).

According to subcellular localization prediction analysis (Supplementary Table 3), the gene XAC0798 encodes an extracellular α -amylase, as also observed in *X. campestris* pv. *campestris* strain 8004²³. This enzyme is classified in subfamily GH13_27, which harbors enzymes that catalyze the endohydrolysis of α -(1 \rightarrow 4)-D-glycosidic linkages in starch, releasing glucose, limit dextrin, maltotriose and maltose²⁴. After the action of this enzyme, oligosaccharides derived from starch can be transported through the TBDTs to the periplasm. In our RNA-seq experiments, we observed the activation of the genes encoding the TBDTs *btuB* (XAC2600), belonging to the Sus loci, and *iroN* (XAC3311) and *btuB* (XAC3444), suggesting that either more than one transporter could be related to starch oligosaccharides uptake or that starch also induces the uptake of other substrates. The genomic context of XAC3311 supports a role in carbohydrate uptake, since it is located near GH2 (XAC3312) and GH97 (XAC3313) encoding genes, but the same is not observed for *btuB* (XAC3444).

At the periplasm, a putative α -glucosidase (XAC2599) might hydrolyze α -glycosidic linkages, releasing glucose from the oligosaccharides. The protein encoded by XAC2599 shows around 51% of sequence identity with the α -glucoside hydrolase *PspAG97A* from

Pseudoalteromonas sp. K8. Different of typical α -glucosidases, *PspAG97A* shows a broad specificity for α -glycosidic linkages (α -1,4-; α -1,2- and α -1,6), with higher catalytic activity on long-chain substrates harboring α -1,6-linkages, releasing α -D-glucose^{25,26}.

The mono and oligosaccharides generated in the periplasm, such as maltose, are probably transported to the cytoplasm through an MFS transporter (XAC2597). Next, α -(1 \rightarrow 4)-D-glucan oligosaccharides might be cyclized by a putative cyclomaltodextrin glucanotransferase (XAC2596, classified in family GH13, not yet assigned to a subfamily), generating cyclodextrins. Also in the cytoplasm, the maltose previously released might than be cleaved into glucose by the putative α -glucosidase (XAC2602). This enzyme shows high sequence identity (91%) with the α -glucosyl transferase XgtA from *Xanthomonas campestris* WU-9701²⁷. XgtA shows both α -glucosidase activity, with high substrate specificity to maltose, and α -glycosylation activity toward alcoholic and phenolic –OH groups, using maltose as an α -glucosyl donor to generate glycoconjugates *in vitro*²⁸. In other words, this enzyme can break down maltose, generating either two glucose molecules, or releasing one glucose and transferring the other to an acceptor molecule, generating glycoconjugates. The physiological relevance of this α -glycosylation activity is still elusive, but a possible role for these glycoconjugates could be in signaling pathways since they seem to be specifically produced during starch processing in *X. citri*.

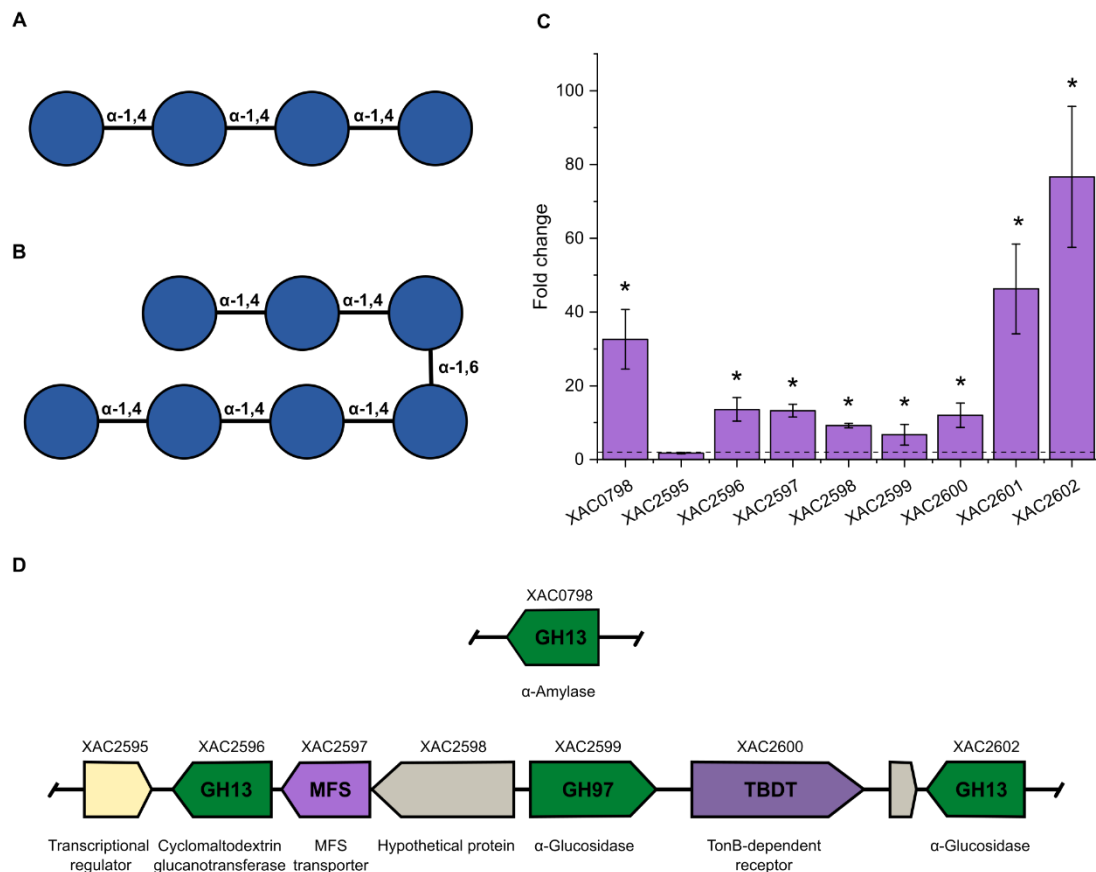


Figure 4. Starch utilization system of *X. citri* strain 306. Representation of starch structure showing **A)** amylose and **B)** amylopectin chains. Blue circles represent glucose residues. **C)** Transcription levels of starch utilization genes in RNA-seq experiments. Data shown as mean \pm SD from four biologically independent experiments ($n = 4$). Threshold: p -adjusted < 0.05 and Fold Change > 2 . **D)** Predicted Starch utilization system (Sus) in *X. citri* genome, showing a transcriptional regulator (yellow), MFS and TBDT transporters (purple), glycoside hydrolases (green) and hypothetical proteins (grey). The hypothetical protein XAC2598 has 780 residues and displays a short motif (163-342) homologous to amylo- α -1,6-glucosidases²⁹, but has not been assigned to a known GH family yet.

Starch inhibits sulfur metabolism and stimulates a sessile state in *X. citri*

GO enrichment analysis for the starch condition showed the suppression of 15 GO categories contrasting with the activation of only 1 category (Hydrolase Activity on glycosyl bonds) (Figure 5A). Over Representation Analysis (ORA) of biochemical pathways revealed that sulfur metabolism is also affected by starch, leading to the suppression of 12 genes related to this pathway, including a putative NADPH-sulfite reductase (Figure 5B and 5C, Supplementary Table 4, Supplementary Data 1). Genes from the upregulated category include three from the Sus loci, which are between those with the highest values of $-\log_{10}P$ and \log_2

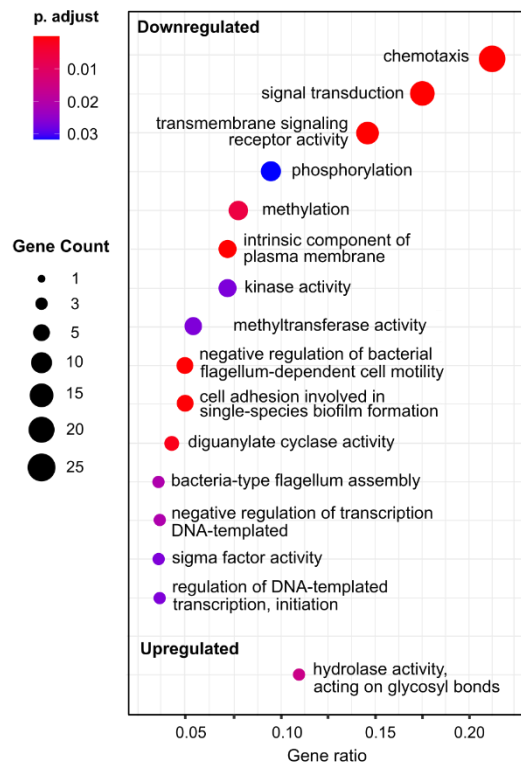
Fold Change (Figure 5C, Supplementary Data 2), indicating that the focus of *X. citri* 306 when exposed to starch is the breakdown of this complex polysaccharide.

Regarding the 15 downregulated categories, those named “Chemotaxis”, “Signal Transduction”, “Transmembrane Signaling Receptor Activity” and “Intrinsic Components of the Plasma Membrane” comprise mainly genes encoding for membrane-embedded methyl-accepting chemotaxis proteins (MCP), chemoreceptors that sense intracellular or environmental signals and direct bacterium movements³⁰ (Supplementary Data 2).

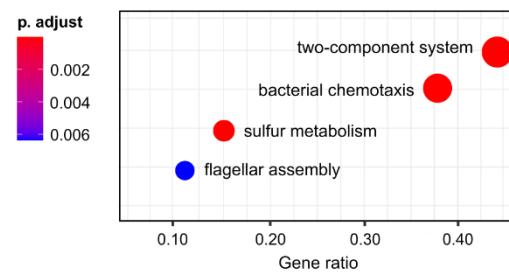
Other downregulated categories represent genes involved with transcriptional regulation, flagellar assembly, and methyl-related activities (Figure 5A, Supplementary Data 2). Downregulated genes related to transcriptional regulation included the one encoding LexA (XAC1739), a repressor of genes involved in the response to DNA damage³¹, the regulators of flagellar genes expression *flgM* (XAC1989, anti- σ factor) and *fliA* (XAC1933, σ factor)³², and the regulators of stress responses and virulence *rseA* (XAC1320, anti- σ factor) and *rpoE1* (XAC1319, σ factor)^{33,34}. Of note, the *lexA* repressor and the pair XAC1319-1320 are also downregulated by cellobiose and other carbohydrates (in the case of *lexA*), indicating they might not be involved in the modulation of starch-specific responses. On the other hand, the flagellar regulators XAC1989 and XAC1933 seem to be repressed only by starch and xyloglucan¹², which correlates with the downregulation of flagellum-related genes specifically in these conditions, including those encoding for the chaperones FlhS (XAC1973), FlgA (XAC1988) and FlgN (XAC1990)^{35, 36,37}.

Several genes encoding proteins with GGDEF and/or EAL domains, which are usually related to the control of intracellular cyclic-di-GMP levels, were suppressed in the starch condition (Figure 5A, Supplementary Data 2). Most of them have been previously observed to be downregulated by xyloglucan in *X. citri*¹², except for the genes XAC0610 and XAC1939. The molecule cyclic-di-GMP is a second messenger recognized as a coordinator of lifestyle changes from motility to sessility and from virulent to less virulent states³⁸. Higher levels of cyclic-di-GMP usually inhibit motility and promote sessility and biofilm formation³⁹. As the starch condition downregulated flagellar assembly and the enzyme responsible for biofilm dispersion in *Xanthomonas* (ManA – XAC1796)^{40,41}, this led us to propose that starch triggers the increase in cyclic-di-GMP levels, favoring a sessile state in *X. citri*.

A) GO enrichment analysis



B) ORA of biochemical pathways



C) Volcano plot

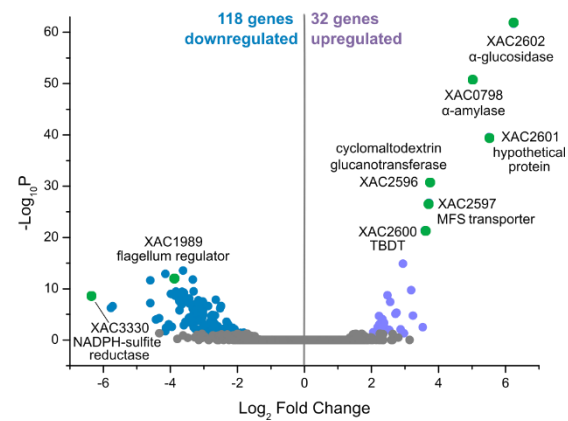


Figure 5. Transcriptional profile of *X. citri* in the presence of starch in minimal medium. A) Gene ontology (GO) enrichment analysis and B) Over Representation Analysis (ORA) of biochemical pathways considering the differentially expressed genes (DEGs) in the minimal medium XVM2m containing starch in comparison to the medium XVM2m containing glucose. Circles size and color represent the counts and BH adjusted p-values, respectively. Gene ratio corresponds to the number of DEGs related to a GO term divided by the total number of annotated DEGs. C) Volcano plot of RNA-seq data showing the DEGs in the presence of starch. In green are indicated some of the DEGs with the higher values of $-\log_{10}P$ and $|\log_2 \text{Fold Change}|$. Thresholds: $p\text{-adjusted} < 0.05$ and $|\log_2 \text{Fold Change}| > 1$.

X. citri displays an unconventional system for galactomannan utilization and a modest transcriptional response to GMOs

Mannans are hemicellulosic polysaccharides typically found in hardwoods and gymnosperms, showing a structural role in the plant cell wall through non-covalent interactions with cellulose^{42,43}. In addition, mannans can also function as a storage carbohydrate in the endosperm walls and vacuoles of seeds^{44,45}. The backbone structure of mannans consists essentially of mannose or a combination of mannose and glucose residues joined by β -1,4 glycosidic bonds, being classified into four types: linear mannan, a mannose homopolymer with β -1,4-glycosidic linkages; glucomannan, a linear and heterogeneous main chain containing both mannose and glucose residues; galactomannan, which has a main chain of mannose residues

and side chains composed by α -1,6-galactose units (Figure 6A); and galactoglucomannan, which has a heterogeneous main chain containing mannose and glucose residues with substituents of α -1,6-galactose⁴⁶.

Regarding the GH-encoding genes found in *X. citri* genome, only XAC3312, a GH2 member of unknown function, was activated in the condition containing galactomannan oligosaccharides (GMOs), according to our RNA-seq data (Figure 6B, Supplementary Data 1). This gene shows low sequence similarity (~24%) with a β -D-mannosidase from *Dictyoglomus thermophilum* H-6-12 (PDB 5N6U)⁴⁷ and was specifically activated by the GMOs condition, indicating that it might encode a β -mannosidase. Close to XAC3312, a gene encoding for a TBDT (XAC3311) was also activated, indicating a role for this transporter in GMOs uptake. Besides XAC3311, the TBDT codified by XAC1769, which belongs to a loci related to xyloglucan depolymerization¹², was also upregulated, supporting the idea that one TBDT can be related to the uptake of more than one type of oligosaccharides in *X. citri*.

Near to the genes XAC3311 and XAC3312, there are other 2 genes that show low sequence identity (< 35%) with an α -galactosidase (XAC3313, assigned in the family GH97) and a carboxylesterase (XAC3315) according to BLAST analysis (Supplementary Table 5). Although these genes were not upregulated in the GMOs condition, their proximity to the loci XAC3311-XAC3312 and their putative enzymatic activities indicate that these genes might also play a role in galactomannan depolymerization in *X. citri*, probably acting in the periplasm (Figure 6C, Supplementary Table 6).

Curiously, no gene classified as endo- β -1,4-mannanase was upregulated in the GMOs condition. In the *X. citri* genome, only 2 genes are classified in GH families harboring this type of activity: XAC1796 and XAC3522. The gene XAC1796 belongs to the family GH5 subfamily 8 and encodes the protein ManA, predicted to be extracellular (Supplementary Table 6). This enzyme is known to play a role in biofilm dispersion in the *Xanthomonas* genus⁴⁰, showing no activity in xanthan gum *in vitro*, but being active on different mannan poly- or oligosaccharides^{41,48}. The other enzyme is encoded by the gene XAC3522, which is classified in the GH113 family. Although this enzyme can cleave mannans *in vitro*⁴⁹, it belongs to a gene cluster dedicated to the synthesis of the Xag-EPS⁵⁰ supporting the concept that the enzyme ManA might be the one responsible for the extracellular galactomannan depolymerization in *X. citri*.

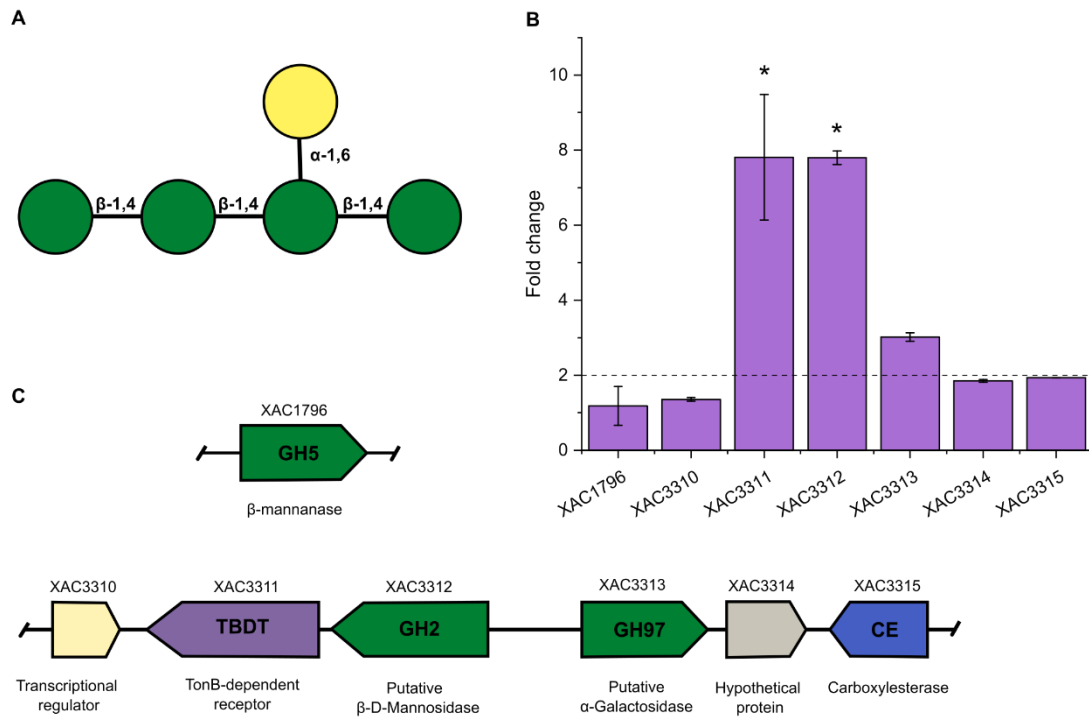


Figure 6. Galactomannan utilization loci of *X. citri* strain 306. **A)** Representation of galactomannan structure. Green and yellow circles represent mannose and galactose residues, respectively. **B)** Transcription levels of galactomannan utilization genes in RNA-seq experiments. Data are shown as mean \pm SD from two biologically independent experiments ($n = 2$). Threshold: p -adjusted < 0.05 and Fold Change > 2 . **C)** Predicted galactomannan utilization system in *X. citri* genome, showing a transcriptional regulator (yellow), a TnB-dependent transporter (purple), glycoside hydrolases (green) and a hypothetical protein (grey). XAC3314 has a DUF239 domain, which is assigned as neprosin⁵¹ with prolyl endopeptidase activity. However, the biological role of this gene is still elusive.

The GMOs condition induced modest changes in the transcriptional profile of *X. citri*, showing only 11 DEGs when compared to the glucose condition (Figure 7). According to GO enrichment analysis, GMOs repressed genes related to protein folding (XAC1585, peptidyl-prolyl cis-trans isomerase), transcriptional regulation (XAC4023, two-component system regulatory protein) and the TonB-like protein encoded by XAC1651 (Figures 7A and 7C, Supplementary Data 2). On the other hand, it activated genes related to GMOs uptake and hydrolysis, as described above, besides genes related to cell redox homeostasis (XAC3878, disulfide isomerase) along with a protein containing a carboxypeptidase regulatory-like domain (XAC3419), classified in the categories “carbohydrate binding” and “carboxypeptidase activity” (Figure 7A and 7C, Supplementary Data 2).

ORA showed only one enriched category containing the transcriptional regulator *phoP* (XAC4023) (Figure 7B). Besides contributing with antimicrobial peptide resistance in

bacteria⁵², PhoP also participates in a global signaling network in *Xanthomonas*, influencing the expression of multiple virulence factors and being necessary for biofilm formation^{53,54}. Interestingly, both GMOs and starch conditions downregulated the expression of *phoP* and its neighbor XAC4024, which is also important for biofilm formation⁵⁵ (Supplementary Data 1). Besides these genes, the only DEG exclusively shared between these two conditions was the upregulated gene XAC3311, which encodes the TBDT from the proposed galactomannan utilization system (Supplementary Data 1). The concomitant downregulation of XAC4024 and upregulation of XAC3311 has also been observed in a *X. citri* $\Delta phoP$ strain⁵³, indicating that PhoP might act as an activator of XAC4024 expression and a repressor of XAC3311 expression. Together, these results indicate that GMOs might negatively affect biofilm formation and stimulates its own uptake by inhibiting *phoP* expression.

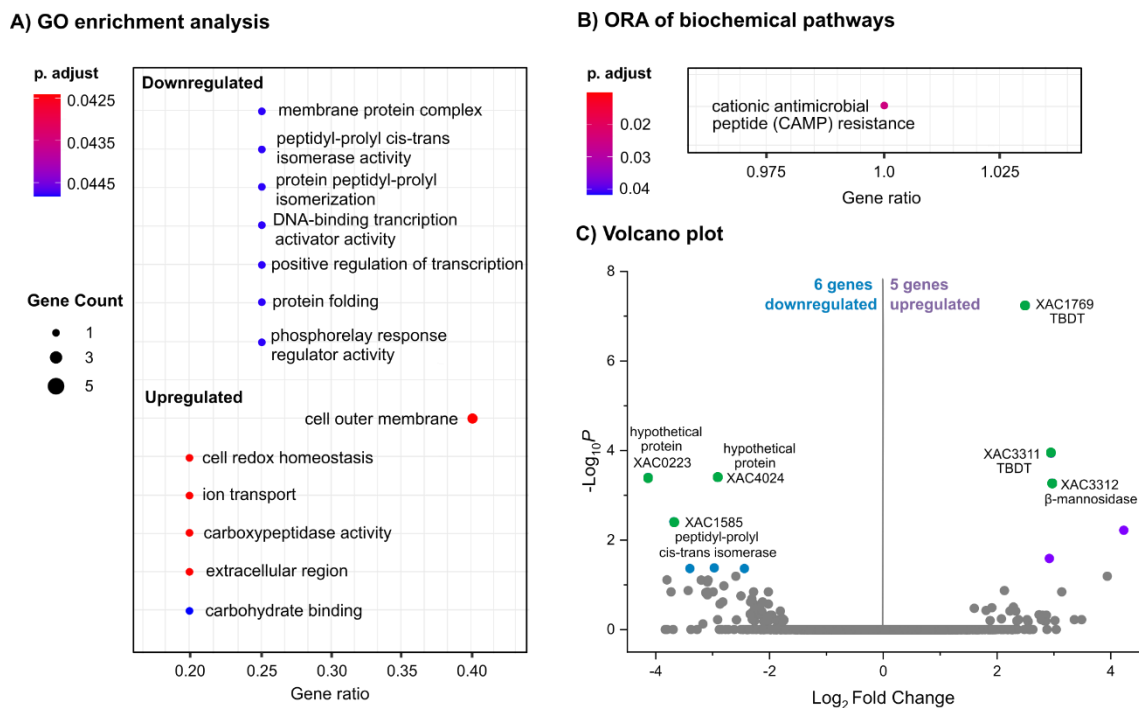


Figure 7. Transcriptional responses of *X. citri* to galactomannan. A) Gene ontology (GO) enrichment analysis and B) Over Representation Analysis (ORA) of biochemical pathways considering the differentially expressed genes (DEGs) in the minimal medium XVM2m containing galactomannan oligosaccharides (GMOs) in comparison to the medium XVM2m containing glucose. Circles size and color represent the counts and BH adjusted p-values, respectively. Gene ratio corresponds to the number of DEGs related to a GO term divided by the total number of annotated DEGs. C) Volcano plot of RNA-seq data showing the DEGs in the presence of GMOs. In green are indicated some of the DEGs with the higher values of $-\log_{10}P$ and $|\log_2 \text{Fold Change}|$. Thresholds: $p\text{-adjusted} < 0.05$ and $|\log_2 \text{Fold Change}| > 1$.

Xylan utilization in *X. citri* likely involves unprecedented GH115 and GH2 enzymes

Xylan is a group of hemicelluloses that share a common backbone of β -1,4-linked xylose residues and may contain different decorations according to the species or tissue, being more abundant in grasses and in dicot secondary cell walls⁴². In dicots, xylans are mainly substituted with α -1,2-linked glucuronic acid (GlcA), 4-O-methyl-glucuronic acid (MeGlcA) and O-acetyl groups composing glucuronoxylans^{56,57} (Figure 8A). In grasses, the xylans may display O-acetyl substituents, α -1,3/ α -1,2-linked arabinofuranosyl (Araf) and α -1,2-linked GlcA/MeGlcA in glucuronoarabinoxylans, or arabinose and O-acetyl substituents in the arabinoxylans, besides showing galactose, ferulic or coumaric acid in small amounts^{56,57} (Figure 8B).

From the possible candidates to encode enzymes related to xylan depolymerization in *X. citri*, only XAC4254 (GH10) and XAC4194 (GH115) were upregulated in the XOs condition (Figure 8C, Supplementary Data 1). XAC4254 encodes a xylanase⁵⁸ and belongs to a cluster that has been associated with xylan utilization in *X. campestris* and *X. citri*^{14,59}. Although not meeting the criteria to be considered differentially expressed, most genes comprising this cluster and other loci showed at least a two-fold increase in expression in the presence of XOs and/or AXOs compared to the glucose condition (Figure 8C). Our data displayed a similar pattern of activation compared to a previous study on *X. citri*, except for the gene XAC4227 (putative alpha-glucuronidase), which was one of the most activated in a previous work⁵⁹, but not in our assays. This difference might be related to the distinct oligosaccharides used in each study (AXOs or XOs here and 4-O-methylglucuronoxylan oligosaccharides in the previous study⁵⁹).

Compared to the study in *X. campestris* pv. *campestris*¹⁴, similarities include the high activation of the glucuronate transporter gene (XAC4255, equivalent to XCC4119). In contrast, the other two genes most activated by xylan in *X. campestris* corresponded to TBDTs (XCC2828, XCC4120), whereas in our study they refer to a xylanase XAC4254 (XCC4118) and to a truncated MFS transporter (XAC4253). These dissimilarities might be likely explained either by the differences in the type of substrates provided in the medium (polymeric¹⁴ vs oligosaccharides in our study) or also to interspecies variations.

Besides XAC4254, *X. citri* encodes other two GH10 enzymes (XAC4249 and XAC4252). XAC4254 is a classical endo-1,4- β -xylanase⁵⁸, which is compatible with a role in the breakdown of xylan polysaccharides at the extracellular medium. Contrasting with a

previous study⁵⁹, plate assays showed that *X. citri* displays extracellular xylanase activity (Supplementary Figure 1). Moreover, studies in *X. campestris* pv. *campestris* and *X. campestris* pv. *vesicatoria* demonstrate that XAC4254 orthologs (XCC4118, 85.7% sequence identity and XCV4360, 95.2% sequence identity)^{14,60} are secreted to the extracellular medium and present in outer membrane vesicles (OMVs), suggesting that XAC4254 might also be found in such locations, despite *in silico* analysis predicts a periplasmic localization (Supplementary Table 7).

In contrast, XAC4249 is an exo-oligoxyylanase specialized in the cleavage of oligosaccharides⁵⁸ and predicted to be periplasmic, according to subcellular localization analysis (Supplementary Table 7). The XAC4249 ortholog in *X. campestris* pv. *vesicatoria* (XCV4355, 97.9% identity) is also a periplasmic protein⁶⁰, supporting a role for XAC4249 in the depolymerization of XOs at the periplasm or at OMVs. The third xylanase encoded by *X. citri* (XAC4252) has not been characterized yet. It is predicted to be localized in the periplasm (Supplementary Table 7), but orthologs in *X. campestris vesicatoria* (XCV4358, 89% amino acid identity)⁶⁰ and *X. oryzae* pv. *oryzae* (XOO4428, 87% amino acid identity) proved to be secreted⁶¹, showing that the biological role and cellular localization of XAC4252 requires further investigation.

In *X. campestris*, a putative α -glucuronidase Agu67A (XAC4227 ortholog) was shown to enhance the extracellular xylanase activity¹⁴, likely by removing glucuronic acid decorations, indicating a similar role in *X. citri*. Other accessory enzymes include the non-reducing-end xylose-releasing exo-oligoxyylanase (GH43, XAC4258)⁶², predicted to be cytoplasmic, a β -xylosidase (GH3, XAC4231)¹² and a putative β -xylosidase/ α -L-arabinofuranosidase (GH43, XAC4230), predicted to be periplasmic (Supplementary Table 7). Interestingly, genes of GH43 putative α -L-arabinofuranosidases (XAC1275 and XAC4183) were not activated in the AXOs condition, indicating that their biological roles might be unrelated to arabinoxylan depolymerization.

Intriguingly, near to the exo-oligoxyylanase gene XAC4249, there is a member of GH2 family (XAC4250) not yet characterized. The family GH2 is multi-specific, including the activities β -galactosidase, β -glucuronidase, β -xylosidase, among others¹¹. Although XAC4250 has been annotated as a putative β -galactosidase, its proximity to a putative glucuronate isomerase gene (XAC4251) indicates a possible role for this GH2 enzyme in the release of β -glucuronic acid from oligosaccharides (Figure 8D). However, no β -linked glucuronic moieties has been reported for xylans so far, and the β -glucuronidases already characterized in the family

GH2 have been mainly associated with the depolymerization of pectins^{63,64}. Thus, future studies will be required to determine the biochemical function of XAC4250 and elucidate whether it plays a role in xylan depolymerization.

In our experiments, the presence of XOs also activated the gene XAC4194, which encodes a hypothetical protein located outside the xylan-utilization cluster previously reported (Supplementary Table 7). This gene is absent in *X. campestris campestris*, but present in other *Xanthomonas* species (Supplementary Table 8). XAC4194 belongs to the GH115 family, for which only the xylan α -1,2-glucuronidase and α -(4-O-methyl)-glucuronidase activities have been reported so far¹¹. The protein encoded by XAC4194 displays 39% sequence identity with the enzyme BoAgu115A from *Bacteroides ovatus* (PDB 4C90), which shows a glucuronoxylan-specific α -glucuronidase activity with strong preference for uronic acids that decorate internal xylose units⁶⁵, indicating that XAC4194 probably display a similar role in *X. citri*. Intriguingly, genes related to starch degradation and uptake (XAC2596, XAC2597 and XAC2599) were also upregulated in the presence of XOs (Supplementary Data 1). As there is no clear evidence for the action of these enzymes in xylan utilization, we hypothesize that this activation might represent a crosstalk between XOs and starch utilization systems.

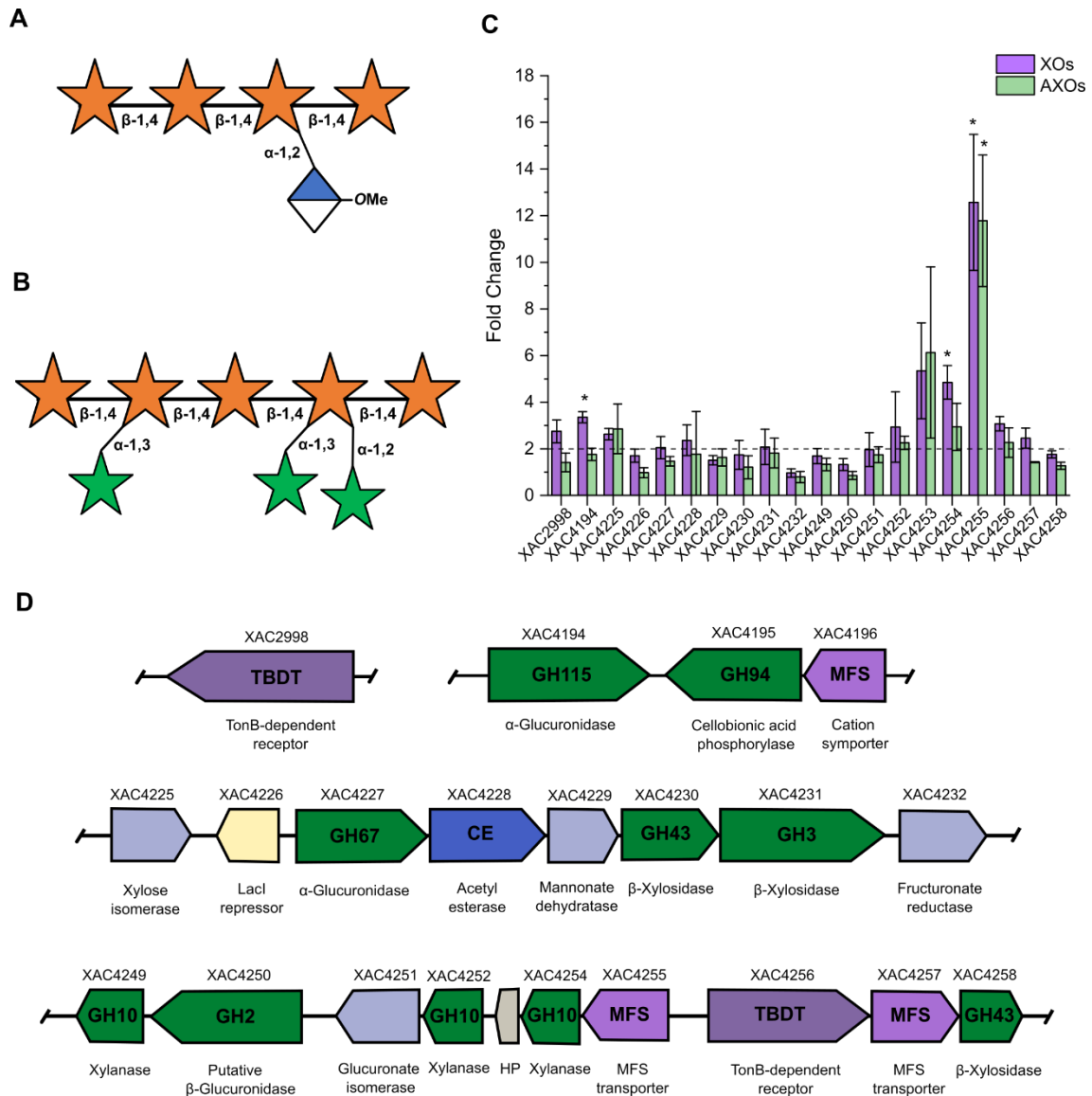


Figure 8. Xylan utilization system of *X. citri* strain 306. Representation of **A)** xylan and **B)** arabinoxylan structures. Orange and green stars represent xylose and arabinose residues, respectively. White and blue rhombus indicate a 4-O-methyl-glucuronic acid (MeGlcA) residue. **C)** Transcription levels of xylan utilization genes in RNA-seq experiments. Data are shown as mean \pm SD from four or three biologically independent experiments ($n = 4$ for XOs and $n = 3$ for AXOs). Threshold: p -adjusted < 0.05 and Fold Change > 2 . **C)** Xylan utilization system in *X. citri* genome, showing a transcriptional regulator (yellow), a TBDT transporter (purple), glycoside hydrolases (green) and a hypothetical protein (HP; grey) according to gene expression data and literature^{14,59}.

Differences in heteroxylan composition are sensed by *X. citri* triggering specific responses with minimal superposition

In our RNA-seq analysis, we evaluated *X. citri* response in the presence of XOs and AXOs and notably observed a different profile of global gene expression in both conditions, with few shared DEGs, indicating that changes in xylan decorations are sensed by *X. citri*

(Figure 9A and B). The only gene simultaneously upregulated in both conditions corresponded to the glucuronate transporter gene XAC4255 from the xylan utilization loci (Figure 9B, Supplementary Data 1). Among the downregulated genes, five were shared between these conditions, corresponding to the genes encoding the transcriptional regulators *mexL* (XAC3461), *lexA* (XAC1739) and *ohrR* (XAC0281), besides the multidrug efflux transporter genes *mexB* (XAC2843) and *mexA* (XAC2844). Of note, MexL regulators have been associated with the repression of multidrug efflux transporter genes in *Pseudomonas aeruginosa*^{66,67}. Despite the distance in the *X. citri* genome, the concomitant repression of *mexL* (XAC3461), and the multidrug efflux transporter components *mexB* (XAC2843) and *mexA* (XAC2844) genes in the XOs and AXOs conditions led us to suggest that MexL may regulate *mexA* and *mexB* expression, but in this case, as a transcriptional activator. MexL belongs to the TetR/AcrR family of transcriptional regulators, which include both repressors and activators, supporting this hypothesis⁶⁸.

Downregulated genes in the presence of XOs, but not of AXOs, included a manganese transport protein (XAC2036) and a hypothetical protein (XAC3927), which were between the most repressed genes (Figure 9C). Exclusively in the AXOs condition, the most downregulated gene was a putative recombinase (XAC2904). For exclusive upregulated genes, the most prominent included hypothetical proteins (XAC3865 and XAC3927 in the AXOs condition and XAC2156 in the XOs condition) (Figure 9C, D).

COG enrichment analysis also evidences the (dis-)similarities between XOs and AXOs-induced responses. The only enriched category shared between these conditions was “DNA binding”, corresponding mainly to the downregulation of transcriptional regulators (Figure 9E and G; Supplementary Data 2). DNA recombination seemed to be suppressed by AXOs, but unaffected by XOs, whereas efflux transport mediated by outer membrane proteins was specifically suppressed by XOs (Figure 9E and G). Upregulated categories were only observed for XOs and refer mostly to genes related xylan or starch utilization, in agreement with ORA analysis (Figure 9E and F). Lack of AXOs-upregulated categories might be due to the fact that 22 out of 34 genes upregulated in this condition are classified as conserved hypothetical proteins and show no biological function assigned so far.

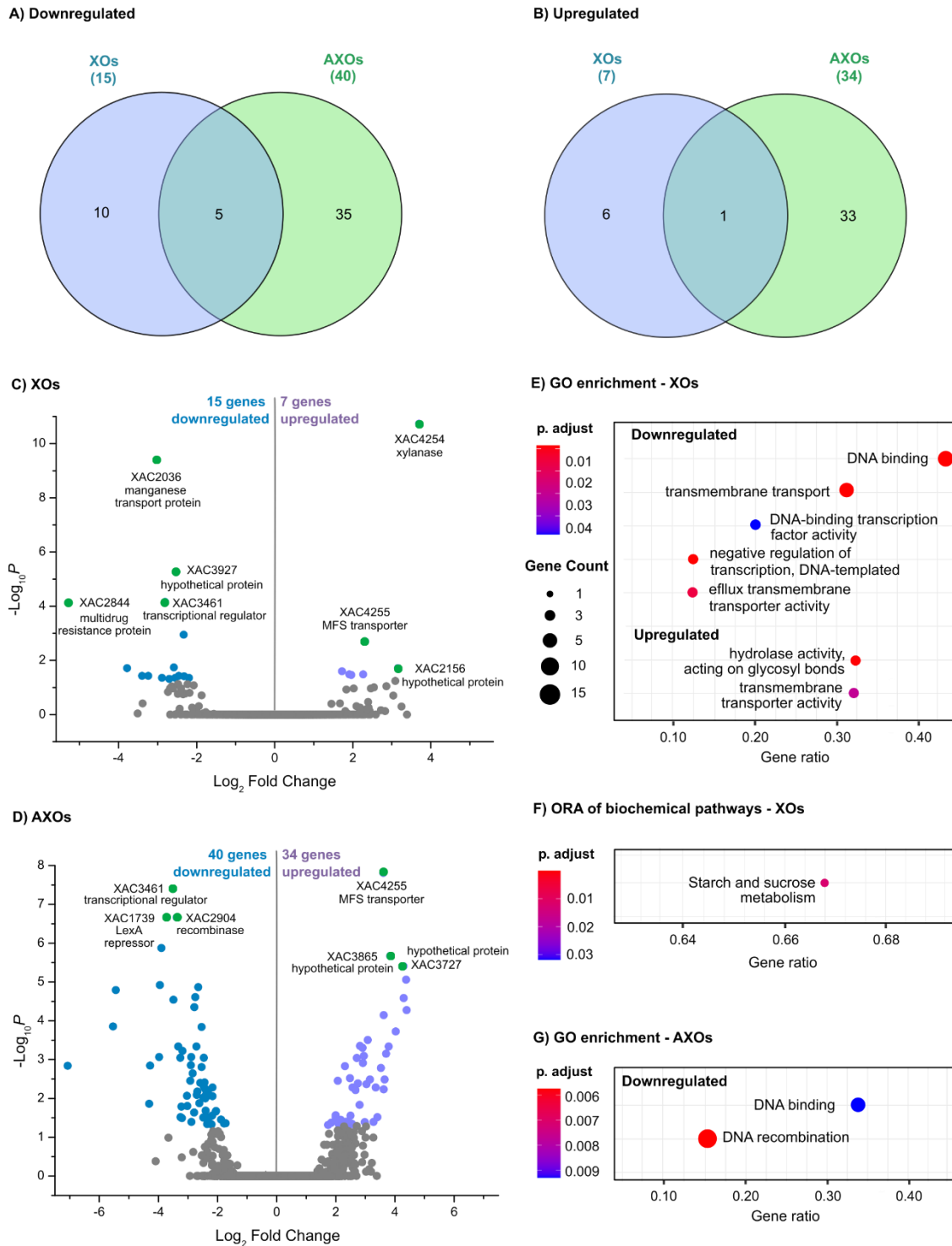


Figure 9. Comparison of the transcriptional profile of *X. citri* in the presence of xylan oligosaccharides (XOs) and arabinoxylan oligosaccharides (AXOs) in minimal medium. Volcano plots of RNA-seq data showing differentially expressed genes (DEGs) in the minimal medium XVM2m containing **A)** XOs or **B)** AXOs in comparison to the medium XVM2m supplemented with glucose. In green are indicated some of the DEGs with the higher values of $-\log_{10} P$ and \log_2 fold change. Genes were considered differentially expressed according to Wald test implemented in DESeq2. p-values were adjusted for multiple tests using the Benjamini-Hochberg (BH) method implemented in DESeq2.

Thresholds: P-adjusted < 0.05 and | log2 Fold Change| > 1. Venn diagrams comparing the distribution of unique and shared **C)** upregulated and **D)** downregulated differentially expressed genes in the minimal medium XVM2m containing XOs or AXOs.

Cellobiose induces the expression of GH5 endoglucanases and several TBDTs for β -glucans utilization in *X. citri*

Cellobiose is a disaccharide composed by D-glucose residues connected by a β -1,4-linkage (Figure 10A). It can be released from cellulose, mixed-linked β -glucans or even from the xyloglucan main chain. Cellulose is one of the main polysaccharides of the plant cell wall matrix, showing a linear chain of β -1,4-D-glucose residues that composes microfibrils, with both crystalline and amorphous regions, that interact with lignin and hemicelluloses⁶⁹. The mixed-linked β -glucans are also composed of D-glucose residues but shows a different pattern of glycosidic linkages (β -1,3; β -1,4 and/or β -1,6) and varies in terms of branching content, molecular weight, and solubility according to the source and type of extraction⁷⁰.

Analysis of *X. citri* genome revealed 5 proteins assigned as endoglucanases from family GH5 (XAC0028, XAC0029, XAC0030, XAC0346, XAC0612), 1 from family GH8 (XAC3516) and 1 from family GH9 (XAC2522). From a previous study, three GH3 enzymes were identified as candidates for cello-oligosaccharides hydrolysis (Supplementary Table 2). Of note, none LPMO or CBH encoding genes were found in the *X. citri* genome, indicating that this bacterium might not be capable to degrade the crystalline portions of cellulose. However, other *Xanthomonas* species presents a CBH encoded by *cbsA*, which is a key factor to allow the infection of vascular systems and might be important to the degradation of cellulose microfibrils in pit membranes of xylem vessels⁷¹.

To find clues about which genes contribute with cellulose and β -glucans degradation in *X. citri*, we analyzed the transcriptional response of *X. citri* to cellobiose, a representative degradation product from these polysaccharides (Supplementary Data 1). Notably, using the glucose condition as a reference, no glycoside hydrolases were upregulated in the presence of cellobiose. Thus, we performed another comparison using RNA-seq data from *X. citri* grown in a rich medium as a reference, based on previous reports showing that the minimal medium XVM2 induce the expression of endoglucanases when compared to a rich medium^{72,73}. This comparison revealed that the genes XAC0028, XAC0030, XAC0346 and XAC0612 were upregulated in the cellobiose condition, suggesting an important role of these endoglucanases in the hydrolysis of cellulose or mixed-linked β -glucans (Figure 10B). XAC0612 has already

been demonstrated to be an endoglucanase⁷⁴, but *in vitro* studies show that XAC0346 does not display endoglucanase activity, likely due to a domain insertion absent in active endoglucanases (data not shown). The biochemical and biological role of XAC0028 and XAC0030 remains so far elusive.

Despite the high adjusted *p*-value, XAC0029 displayed an average fold change near 4 in the cellobiose condition compared to the rich medium, indicating that it might also be involved in β -glucan depolymerization (Figure 10C). For the GH8 and GH9 encoding genes, no activation was observed, suggesting they are not involved in β -glucan breakdown. The genomic context of GH8 gene indicates a role for this already characterized endoglucanase⁷⁵ in bacterial cellulose remodeling, but for the GH9 protein it is still elusive whether or not it contributes to β -glucan depolymerization. Regarding the GH3 β -glucosidases, none of them were upregulated in the cellobiose condition, although previous studies have demonstrated they display β -glucosidase activity¹². In comparison to rich medium, XAC3869 was the only one presenting a positive fold change, suggesting a major role in cellobiose hydrolysis (Figure 10C).

Interestingly, GH-encoding genes with predicted activities unrelated to the cleavage of β -glucans were upregulated in the cellobiose vs rich medium condition, including a hypothetical protein from family GH15 (XAC4082), the α -glucosidase from family GH97 of *X. citri* Sus loci (XAC2599) and the gene XAC3254 encoding a putative glycogen debranching enzyme. The activation of the last two genes might be a response to low carbon and energy availability, since they encode enzymes with a possible role in bacterial glycogen hydrolysis⁷⁶.

All identified endoglucanases display signal peptides (Supplementary Table 9) and previous secretome and proteome studies indicate that XAC0028, XAC0029 and XAC0612 encode extracellular proteins⁷⁷, whereas XAC0030 encodes a periplasmic enzyme⁷⁸. The protein encoded by XAC2522 has a lipoprotein signal peptide, which suggests an anchoring of this protein in the membrane (Supplementary Table 9). Although *in silico* analysis indicates a periplasmic localization for XAC2522, this enzyme was previously reported to contribute to the extracellular activity of *X. citri* on the polysaccharide xyloglucan¹², supporting its anchoring in the outer leaflet of the outer membrane, in a similar way as previously observed for a GH9 of *Bacteroides ovatus*⁷⁹. The GH3 β -glucosidase XAC3869 is predicted to be in the cytoplasm according to sequence-based prediction of subcellular localization analysis (Supplementary Table 9).

To find clues about genes related to cellobiose uptake, we went back to the comparisons of cellobiose vs glucose condition. This analysis revealed the activation of 5 putative TBDTs (XAC0291, XAC0852, XAC3201, XAC3418 and XAC3613), some of which might be possibly involved with cello-oligosaccharides uptake to the periplasm (Supplementary Data 1). When comparing the expression of the TBDT genes in the other growth conditions, we noticed that they were significantly activated only in the presence of cellobiose, besides being already reported to be activated in the presence of xyloglucan-oligosaccharides¹², further supporting their possible role in cello-oligosaccharides uptake in *X. citri*.

According to these analyses, we propose that the degradation of cellulose and mixed-linked β -glucans in *X. citri* starts with the action of XAC0028, XAC0029 and XAC0612 endoglucanases in the extracellular medium and might be assisted by the GH9 endoglucanase anchored in the outer membrane (XAC2522), although this hypothesis requires further validation. After the uptake of cello-oligosaccharides, possibly by one or more of the upregulated TBDTs (XAC0291, XAC0852, XAC3201, XAC3418 and XAC3613), the endoglucanase XAC0030 might break cello-oligosaccharides up to cellobiose at the periplasm. At the cytoplasm, the β -glucosidase XAC3869 might complete the depolymerization process converting the remaining cello-oligosaccharides into glucose.

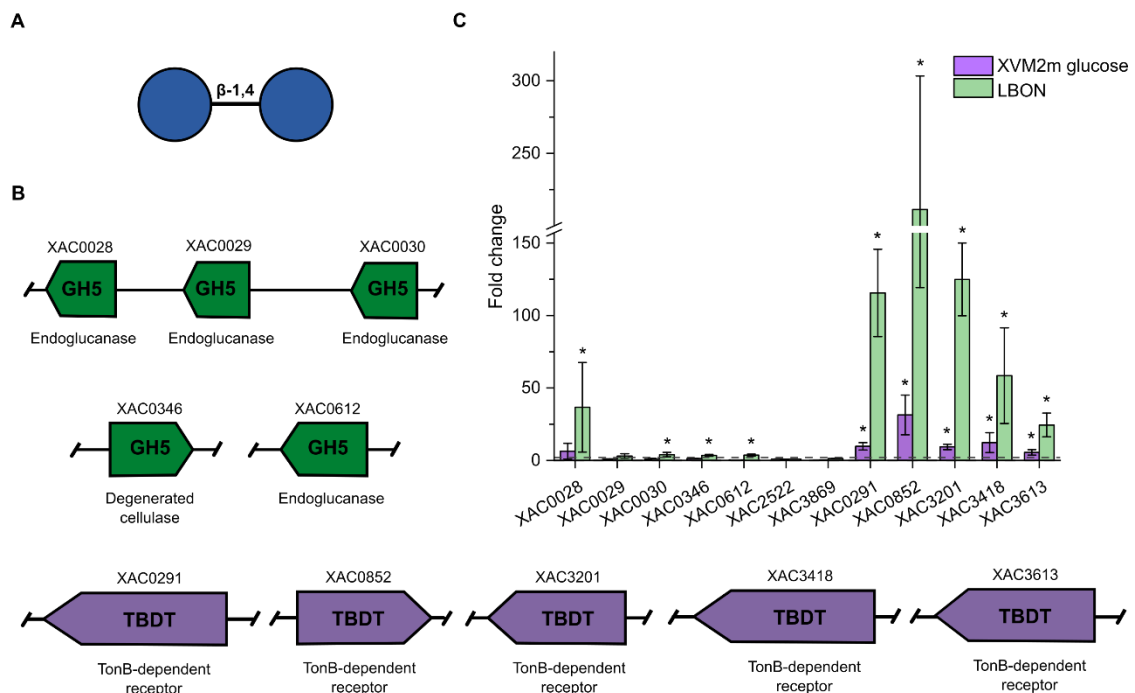


Figure 10. Beta-glucans utilization loci of *X. citri* strain 306. A) Representation of cellobiose structure. Blue circles represent glucose residues. B) Transcription levels of beta-glucans utilization

genes in RNA-seq experiments. Data are shown as mean \pm SD from four biologically independent experiments (n = 4). Threshold: p-adjusted < 0.05 and Fold Change > 2. C) Predicted beta-glucan utilization system in *X. citri* genome, showing a transcriptional regulator (yellow), a TBDT transporter (purple), glycoside hydrolases (green) and a hypothetical protein (grey).

Cellobiose inhibits transcriptional regulators expression and induce motility and sulfur metabolism

The GO enrichment analysis for the cellobiose condition, using the minimal medium XVM2m with glucose as a reference, showed the enrichment of 14 categories of which 9 were activated and 5 were suppressed (Figure 11A). Notably, the downregulated category “DNA binding” contains 26 genes related to transcription or recombination functions which are also partially distributed in the other enriched categories such as “Regulation of transcription” and “transcription factor activity”. Transcriptional regulators from different families (AraC, MarR, tetR, LysR and gntR) were downregulated, suggesting the modulation of several biological functions in the presence of cellobiose. Interestingly, three LysR regulators were exclusively downregulated in the cellobiose condition (XAC0255, XAC3459 and XAC2718), indicating that they might play a role in the regulation of cellobiose-specific responses. Among the genes assigned in the “DNA recombination” category are 4 resolvases (XACb0071, XACa0029, XACa0038, XACb0072), a recombinase from phage-integrase family (XAC2904) and the gene encoding the RecA protein (XAC1740).

Among the upregulated categories, “Plasma membrane” showed the higher number of genes, including six ABC transporters, some of them related to sulfur/sulfonates transport (XAC1018, XAC1019 and XAC3197)⁸⁰, the pre-pilin like leader sequence *fimT* (XAC2669), the flagellar protein *fliL* (XAC1948) and the gene encoding the cytochrome C oxidase subunit III (XAC3884). The category “Protein secretion by the type II secretion system (T2SS)” includes the genes *pilE* (XAC2664), *fimA* (XAC3241) and *fimT*, which are actually related to the type IV pili (T4P)⁸¹ biogenesis and might have been classified as part of the T2SS because of the high similarity of these systems⁸².

The categories “Monooxygenase activity” and “Oxidoreductase activity” include the genes XAC0854, XAC0855 and XAC3200, that although not yet characterized, are probably involved in the metabolism of alkanesulfonates utilization due to their proximity to ABC transporters that have been related to the uptake of these compounds⁸⁰. Furthermore, the category “Bacterial-type flagellum-dependent cell motility” were also enriched, containing the genes *flgL* (XAC1976), *flgG* (XAC1981), *flgE* (XAC1983), as also *fliL*, suggesting an increase

in motility in the presence of cellobiose. Next, the categories “Carboxypeptidase activity”, “Carbohydrate binding”, “Cell outer membrane” and “Ion transport” include all the activated TBDTs in the presence of cellobiose (XAC0291, XAC0852, XAC3201, XAC3418 and XAC3613), supporting a possible role of some of these transporters in the uptake of cellobiose.

In the ORA analysis, the enriched categories are in agreement with the profile observed in the GO analysis, showing the metabolism of sulfur and selenocompounds, as well as flagellar assembly and quorum sensing (Figure 11B). In the volcano plot analysis, the genes with the higher values of $-\log_{10} P$ and $|\log_2 \text{Fold Change}|$ are also representative of the categories observed in enrichment analyses, highlighting the downregulation of the transcriptional regulators *lexA* (XAC1739) and XAC3363, and the activation of the ABC transporter XAC0856, the TBDT encoded by XAC0852 and the pilin XAC3241 (Figure 11C).

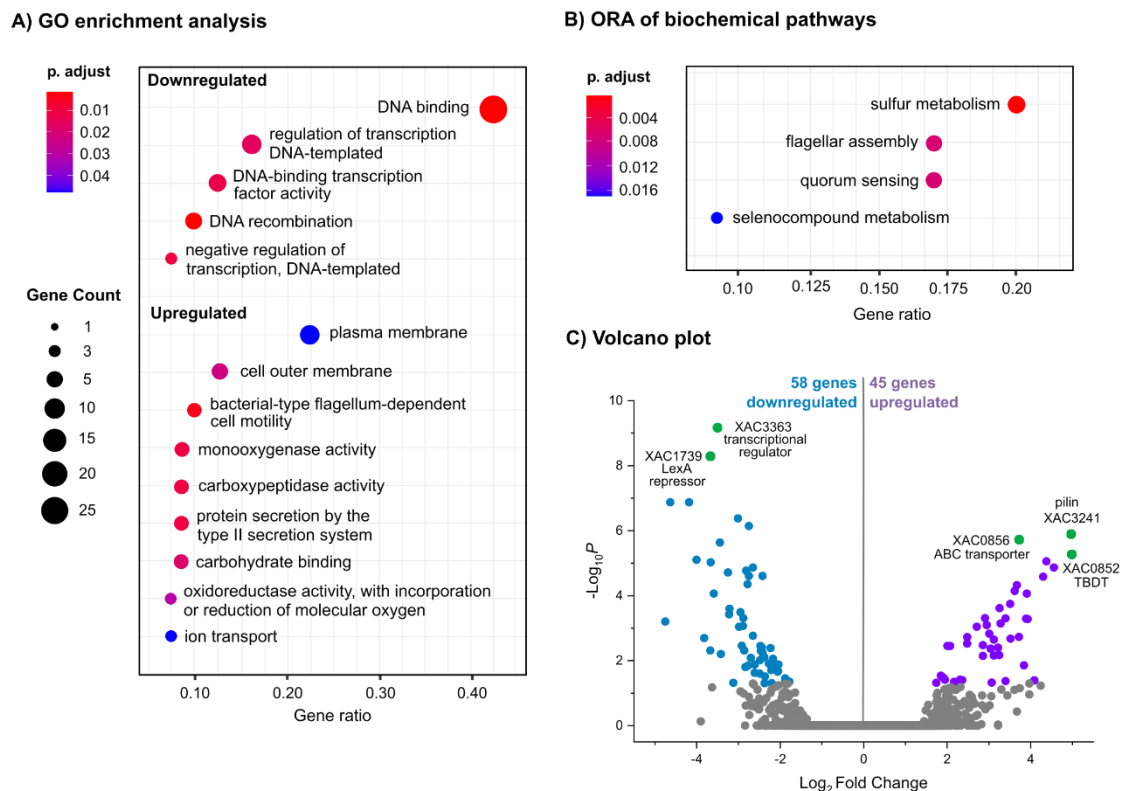


Figure 11. Transcriptional responses of *X. citri* to cellobiose. A) Gene ontology (GO) enrichment analysis and B) Over Representation Analysis (ORA) of biochemical pathways considering the differentially expressed genes (DEGs) in the minimal medium XVM2m containing cellobiose in comparison to the medium XVM2m containing glucose. Circles size and color represent the counts and BH adjusted p-values, respectively. Gene ratio corresponds to the number of DEGs related to a GO term divided by the total number of annotated DEGs. C) Volcano plot of RNA-seq data showing the DEGs

in the presence of cellobiose. In green are indicated some of the DEGs with the higher values of $-\log_{10} P$ and $|\log_2 \text{Fold Change}|$. Thresholds: $p\text{-adjusted} < 0.05$ and $|\log_2 \text{Fold Change}| > 1$.

Discussion

Signaling molecules are primarily responsible for dictating behavioral responses in bacteria, either by directing their movements across the environment, indicating the need to activate protective mechanisms against stress or even by showing them the possibility of conquering new niches. Similarly to what occurs with animal pathogenic bacteria⁸⁵, here we show that host-cell-derived carbohydrates also function as spatial guiders for phytopathogenic bacteria, indicating where they are within the host and stimulating appropriate responses to each niche experienced during the infection process.

In our previous study, we showed that xyloglucan, a hemicellulose present in the plant cell wall, inhibits motility in *X. citri* 306 while activates the expression of genes related to xanthan gum biosynthesis and type-III secretion, which are crucial in the early stages of infection^{12,86}. Here we show that other hemicelluloses might also indicate to *X. citri* 306 the progress of plant cell wall degradation, modulating the bacteria transcriptional response according to their relative abundance in the infection site. Later, increased levels of cellobiose, the last product whose origin from plant cell wall polysaccharides can be identified, may indicate to *X. citri* that the process of plant cell wall degradation is ending, stimulating a transition to a more motile state. This higher motility induced by cellobiose might be important for *X. citri* moving either to attack another site of the plant cell wall or to enter into the intracellular space. The presence of a *X. citri* pv. *citri* strain within damaged plant cells has been observed for the virulent strain T in a previous work⁸⁷, supporting the hypothesis that these bacteria might have access to the cytoplasm and consequently to the starch granules during the host infection. When *X. citri* 306 is exposed to starch, motility and biofilm dispersion are inhibited, favoring cell growth as supported by comparative growth analysis in the carbohydrates evaluated in this study (Supplementary Figure 2).

To use the complex plant polysaccharides as sources of carbon, energy and stimuli, *X. citri* 306 has diverse gene sets dedicated to depolymerize, uptake and metabolize these biopolymers. Based on our results, we identified unprecedented loci for starch (XAC2595-XAC2602) and galactomannan (XAC3310- XAC3315) utilization and revealed a putative GH115 α -glucuronidase (XAC4194) likely involved in xylan depolymerization. Moreover, we also highlighted some knowledge gaps on the biochemical characterization of *Xanthomonas*

spp. CAZyme, which may guide future studies aiming to a deeper understanding of polysaccharides utilization systems in this bacterial genus. Notably, we observed the activation of some TBDTs by more than one type of carbohydrate, suggesting they might play a role in the uptake of oligosaccharides from different sources. Similarly, the activation of enzymes not related with the degradation of the carbohydrate present in the growth condition (e.g. activation of the starch-related genes XAC2596 and XAC2599 in the presence of XOs) indicate that a crosstalk of signals derived from different plant polysaccharides might occur during the infection process.

To the best of our knowledge, this study is the first to show the genome-wide transcriptional response of a phytopathogenic bacterium to a diverse set of complex carbohydrates representative of host cell polysaccharides *in vitro*, giving insights on their specific contribution in modulating bacterial behavior during host colonization. In this regard, we provide a framework for future studies aiming to understand to what extent our findings are conserved in other *Xanthomonas* species or even in other plant pathogens and which molecular mechanisms plant pathogens evolved to use, sense, and adapt their responses to the chemical diversity of plant-derived polysaccharides.

Methods

Growth curves analysis

For growth curve analysis, *Xanthomonas citri* pv. *citri* 306 strain was cultured in LBON medium (1% m/v bacto peptone and 0.5% m/v yeast extract) containing 100 $\mu\text{g} \cdot \text{mL}^{-1}$ ampicillin at 30 °C and 200 rpm until mid-exponential phase. Then, the cultures were centrifuged for 5 minutes at 6000 g and the harvested cells were washed once with the modified minimal medium XVM2¹⁷ (XVM2m, without sucrose and fructose) supplemented with different carbohydrates, and transferred to a 20 mL culture medium for an initial $\text{OD}_{600 \text{ nm}} = 0.01$. Growth analyses were performed at 30 °C and 200 rpm, being monitored through $\text{OD}_{600 \text{ nm}}$ readings at times 0h, 8h, 14h, 20h, 26h, 32h, 36h, 44h and 52h in triplicate (Supplementary Figure 2).

Hydrolysates preparation

As insufficient bacterial growth for RNA extraction was observed for *X. citri* in XVM2m medium containing the polysaccharides β -glucan, xylan, arabinoxylan and galactomannan, different hydrolysates were prepared using these corresponding

polysaccharides pre-digested using purified recombinant endo-enzymes from *X. citri* (with exception of β -glucan, which was substituted for cellobiose from Sigma Aldrich). For each hydrolysate, the XVM2m medium was prepared with 5 mg. mL⁻¹ of the corresponding polysaccharide followed by the addition of the enzyme and incubation in the specific conditions for enzymatic hydrolysis (Supplementary Table 10) to generate oligosaccharides for the growth conditions. All the reactions were stopped by heating at 80 °C for 15 min and then filter sterilized using 0.22 μ m syringe filters (MILLEX®).

Cloning and protein heterologous expression

The XAC3522 nucleotide sequence (Genbank AAM38365.1) was amplified from the genomic DNA of *X. citri* pv. *citri* 306 strain using standard methods (primer forward - GAATTCCATATGGCCGCGCCACCTGGATG and reverse - CTCGAGTTAGCGCGCAAGTACCTGCTGCG). PCR-amplified gene fragments were cloned into the pET28a vector and confirmed by Sanger sequencing.

Next, the mannanase encoded by XAC3522 was expressed in *Escherichia coli* strain BL21(DE3) using Terrific Broth medium (1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.4% (w/v) glycerol, 0.017 M sodium phosphate monobasic and 0.072 M sodium phosphate dibasic) supplemented with kanamycin (50 μ g. mL⁻¹). The cultures were incubated at 37 °C and 200 rpm until the OD_{600nm} had reached ~1.0. Then, the cultures were transferred to the temperature of 20 °C followed by the addition of isopropyl β -d-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induction time was of 16 hours at 20 °C and 200 rpm. Cells were then harvested by centrifugation and the cell pellets were kept at – 20 °C. The xylanase encoded by XAC4254 was expressed as previously reported⁴¹.

Protein purification

The cell pellets from XAC3522 were collected, resuspended in lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 5 mM imidazole and 0.03 mg. mL⁻¹ of DNase) containing a SIGMAFAST™ Protease Inhibitor Cocktail tablet and then disrupted by sonication. After centrifugation at 30000 g and 4 °C for 40 minutes, the soluble extract was filtered and applied into a 5-mL HiTrap Chelating HP column (GE Healthcare), charged with Ni²⁺, pre-equilibrated with the affinity buffer and coupled to an ÄKTA purifier (GE Healthcare). The protein was eluted using a non-linear gradient of imidazole (0-0.5 M) at a flow rate of 2 mL. min⁻¹. Eluted fractions were analyzed by SDS-PAGE and the purest were concentrated by filtration and submitted to a size-exclusion chromatography (SEC) in a HiLoad 16/600 Superdex 200 column

pre-equilibrated with 20 mM sodium phosphate and 150 mM NaCl, pH 7.4. SEC was performed at a flow rate of 1 mL.min⁻¹ using an ÄKTA purifier system (GE Healthcare). The XAC4254 enzyme was purified as previously described⁴¹.

RNA extraction and cDNA library preparation

Total RNA samples were extracted from 15 mL *X. citri* cultures at mid-exponential phase using the TRIzol/chloroform protocol⁸⁸. Genomic DNA was removed by treatment with DNaseI (Invitrogen™) and then the samples were treated with RNaseOUT (Invitrogen™), followed by purification with the RNeasy Mini Kit (Qiagen), according to the manufacturer's recommendations. The absorbance analyzes were performed in a NanoDrop spectrophotometer (Thermo Scientific) and the integrity of the samples evaluated in the Agilent 2100 Bioanalyzer (Agilent Technologies). Prior to the tests, the samples were quantified on a Qubit® 2.0 Fluorometer using the RNA BR assay kit (Life Technologies).

For cDNA library preparation, only RNA samples free of genomic DNA contamination and with values of RIN (RNA Integrity Number) greater than 7 were used. Then, 2-2.5 µg of RNA was used for depletion of rRNA using the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria - Epicenter Biotechnologies). The preparation of the cDNA libraries was performed using the TruSeq Stranded mRNA kit (Illumina Inc.) according to the manufacturer's protocol. Samples quality was accessed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and libraries were quantified via qPCR using the KAPA Library Quantification Kit (Illumina). Paired-end sequencing (100 bp) of the constructed libraries was carried out on an Illumina HiSeq 2500 system at LNBR (CNPq, Campinas, Brazil).

RNA sequencing and analysis

The raw sequencing reads were processed to remove the low-quality sequences and adapters, then filtered for removal of ribosomal RNA using the software SorteMeRNA 2.1⁸⁹. The sequences were then mapped to the *Xanthomonas citri* pv. *citri* 306 strain genome⁹⁰ using the Bowtie2 program⁹¹. The Rsubread software was then used to count reads mapped to *X. citri* transcripts⁹². All reads signed to coding genes were used for differential expression analysis by calculating the TPM values (transcribed per million reads) with the R DESeq2 package version 1.18.1⁹³. The resulting values were log₂ transformed and t-test was performed on these expression values to compare differential gene expression of *X. citri* under the growth condition

XVM2m containing different carbohydrates in comparison to XVM2m containing glucose or LBON medium.

Genes were considered differentially expressed according to Wald test and p-values were adjusted for multiple tests using the Benjamini-Hochberg (BH) method implemented in DESeq2 package, using $|\log_2\text{fold change}| \geq 1$ and a $p\text{-adjusted} \leq 0.05$ as a threshold. Analyses of variance using the PCA method (Principal Component Analysis) and hierarchical clustering based on the correlation of Pearson were also performed to examine data quality and comparability. Non-concordant replicas were removed from the analyses and disregarded to determine differentially expressed genes (DEG).

The Carbohydrate-Active enZymes Database (CAZy)¹¹ was used as a reference for the identification of the DEGs classified in CAZy families. The Gene ontology (GO) enrichment analysis was performed using the clusterProfiler 3.14.3R/Bioconductor package⁹⁴ and the categories were considered enriched based on hypergeometric test, implemented in the enrich function of the package. Over Representation Analysis (ORA) of biochemical pathways was also performed using the clusterProfiler 3.14.3R/Bioconductor package⁹⁴. The analysis of co-expression of the different evaluated conditions was performed using the CEMiTool package²⁰.

Plate assay for carbohydrate degradation

Xanthomonas citri pv. *citri* strain 306 and *Escherichia coli* DH5 α were grown overnight at 30°C 200 rpm, in LBON and LB medium (10 g. L⁻¹ bacto tryptone, 5 g. L⁻¹ yeast extract and 10 g. L⁻¹ NaCl), respectively. Optical density of each culture was standardized to 0.4 and 0.5 μ L each were inoculated, in triplicate, on solid NYG (5 g. L⁻¹ peptone, 3 g. L⁻¹ yeast extract, 20 g. L⁻¹ glycerol and 12 g. L⁻¹ agar), supplemented with 0.5% of wheat arabinoxylan low viscosity (Megazyme) or xylan from beechwood (Sigma Aldrich). Three plates of each polysaccharide were kept at 28 °C for 24, 48 and 72 hours. At each time, plates were photographed, followed by colonies removal with water and stained with a 5 mg. mL⁻¹ Congo red solution for 30 minutes. Carbohydrate degradation zones were observed as clear zones after successive washes with 1 M NaCl solution. *Escherichia coli* DH5 α were grown in the same conditions as a negative control.

Authors' contributions

I.M.B., D.A.A.P., G.F.P., P.O.G. and M.T.M. conceived and designed the experiments. I.M.B. and D.A.A.P. performed the experiments. I.M.B., D.A.A.P., G.F.P. and P.O.G. performed RNA-seq analyses. J.M.J performed enrichment and co-expression analyses. J.A.D. performed plate assays. I.M.B. analyzed the results and wrote the manuscript. P.O.G. and M.T.M. coordinated the work, analyzed the results and wrote the manuscript.

References

1. Cameron, E. A. & Sperandio, V. Frenemies: Signaling and nutritional integration in pathogen-microbiota-host interactions. *Cell Host Microbe* **18**, 275–284 (2015).
2. Fatima, U. & Senthil-Kumar, M. Plant and pathogen nutrient acquisition strategies. *Front. Plant Sci.* **6**, 1–12 (2015).
3. Houston, K., Tucker, M. R., Chowdhury, J., Shirley, N. & Little, A. The plant cell wall: A complex and dynamic structure as revealed by the responses of genes under stress conditions. *Front. Plant Sci.* **7**, 1–18 (2016).
4. He, C. *et al.* Cytochemical localization of polysaccharides in *Dendrobium officinale* and the involvement of *DoCSLA6* in the synthesis of mannan polysaccharides. *Front. Plant Sci.* **8**, 1–11 (2017).
5. Meier, H. & Reid, J. S. G. Reserve polysaccharides other than starch in higher plants. In: *Plant Carbohydrates I* (eds. Loewus, F. A. & Tanner, W.), 418–471 (Springer Berlin Heidelberg, 1982). doi:10.1007/978-3-642-68275-9_11.
6. Schwalm, N. D. & Groisman, E. A. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. *Trends Microbiol.* **25**, 1005–1015 (2017).
7. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, 1–15 (2017).
8. Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics* **31**, 647–655 (2015).
9. Ausland, C. *et al.* dbCAN-PUL: A database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res.* **49**, D523–D528 (2021).
10. Amicucci, M. J. *et al.* Strategy for structural elucidation of polysaccharides: elucidation of a maize mucilage that harbors diazotrophic bacteria. *Anal. Chem.* **91**, 7254–7265 (2019).
11. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
12. Vieira, P. S. *et al.* Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* **12**, 1–15 (2021).
13. Blanvillain, S. *et al.* Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. *PLoS One* **2**, (2007).
14. Déjean, G. *et al.* The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv. *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytol.* **198**, 899–915 (2013).
15. Boulanger, A. *et al.* The plant pathogen *Xanthomonas campestris* pv. *campestris* exploits N-Acetylglucosamine during infection. *MBio* **5**, 1–13 (2014).

16. Dupoirson, S. *et al.* The N-Glycan cluster from *Xanthomonas campestris* pv. *campestris*: A toolbox for sequential plant N-Glycan processing a toolbox for sequential plant N-Glycan processing. *J. Biol. Chem.* **290**, 6022–6036 (2015).
17. Wengelnik, K., Marie, C., Russel, M. & Bonas, U. Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction. *J. Bacteriol.* **178**, 1061–1069 (1996).
18. Heberle, H., Meirelles, V. G., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 1–7 (2015).
19. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**, 14863 LP – 14868 (1998).
20. Russo, P. S. T. *et al.* CEMiTool: A Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* **19**, 1–13 (2018).
21. Bertoft, E. Understanding starch structure: Recent progress. *Agronomy* **7**, (2017).
22. Konur, O. Glycoscience: The current state of the research. In: Marine Glycobiology: Principles and Applications (1st ed.). *CRC Press*. (2016). doi:10.1201/9781315371399-3.
23. Lin, Y. *et al.* Extracellular amylase is required for full virulence and regulated by the global posttranscriptional regulator RsmA in *Xanthomonas campestris* pathovar *campestris*. *Phytopathology* **111**, 1104–1113 (2021).
24. Laia, M. L. *et al.* New genes of *Xanthomonas citri* subsp. *citri* involved in pathogenesis and adaptation revealed by a transposon-based mutant library. *BMC Microbiol.* **9**, 1–17 (2009).
25. He, C. *et al.* Structures of PspAG97A α -glucoside hydrolase reveal a novel mechanism for chloride induced activation. *J. Struct. Biol.* **196**, 426–436 (2016).
26. Li, W. *et al.* PspAG97A: A halophilic α -glucoside hydrolase with wide substrate specificity from glycoside hydrolase family 97. *J. Microbiol. Biotechnol.* **26**, 1933–1942 (2016).
27. Watanabe, R., Arimura, Y., Ishii, Y. & Kirimura, K. Crystal structure of α -glucosyl transfer enzyme XgtA from *Xanthomonas campestris* WU-9701. *Biochem. Biophys. Res. Commun.* **526**, 580–585 (2020).
28. Sato, T. *et al.* Purification, characterization, and gene identification of an α -glucosyl transfer enzyme, a novel type α -glucosidase from *Xanthomonas campestris* WU-9701. *J. Mol. Catal. B Enzym.* **80**, 20–27 (2012).
29. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
30. Salah Ud-Din, A. I. M. & Roujeinikova, A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell. Mol. Life Sci.* **74**, 3293–3303 (2017).
31. Yang, Y. C., Yang, M. K., Kuo, T. T. & Tu, J. Structural and functional characterization of the *lexA* gene of *Xanthomonas campestris* pathovar *citri*. *Mol. Gen. Genet.* **265**, 316–326 (2001).
32. Yang, T. C., Leu, Y. W., Chang-Chien, H. C. & Hu, R. M. Flagellar Biogenesis of *Xanthomonas campestris* requires the alternative sigma factors RpoN2 and FlhA and is temporally regulated by FlhA, FlhB, and FlgM. *J. Bacteriol.* **191**, 2266–2275 (2009).
33. Bordes, P. *et al.* Insights into the extracytoplasmic stress response of *Xanthomonas campestris* pv. *campestris*: role and regulation of σ E-dependent activity. *J. Bacteriol.* **193**, 246–264 (2011).
34. Yang, L. Y. *et al.* Systematic functional analysis of sigma (s) factors in the phytopathogen *Xanthomonas campestris* reveals novel roles in the regulation of virulence and viability. *Front.*

- Microbiol.* **9**, (2018).
35. Auvray, F., Thomas, J., Fraser, G. M. & Hughes, C. Flagellin polymerisation control by a cytosolic export chaperone. *J. Mol. Biol.* **308**, 221–229 (2001).
 36. Matsunami, H., Yoon, Y. H., Meshcheryakov, V. A., Namba, K. & Samatey, F. A. Structural flexibility of the periplasmic protein, FlgA, regulates flagellar P-ring assembly in *Salmonella enterica*. *Sci. Rep.* **6**, 1–3 (2016).
 37. Khater, L. *et al.* Identification of the flagellar chaperone FlgN in the phytopathogen *Xanthomonas axonopodis* pathovar citri by its interaction with hook-associated FlgK. *Arch. Microbiol.* **188**, 243–250 (2007).
 38. Römling, U., Galperin, M. Y. & Gomelsky, M. Cyclic di-GMP: the First 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52 (2013).
 39. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–273 (2009).
 40. Dow, J. M. *et al.* Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10995–11000 (2003).
 41. Hsiao, Y. M., Liu, Y. F., Fang, M. C. & Tseng, Y. H. Transcriptional regulation and molecular characterization of the *manA* gene encoding the biofilm dispersing enzyme mannan endo-1, 4- β -mannosidase in *Xanthomonas campestris*. *J. Agric. Food Chem.* **58**, 1653–1663 (2010).
 42. Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289 (2010).
 43. Marcus, S. E. *et al.* Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J.* **64**, 191–203 (2010).
 44. Rodríguez-Gacio, M. del C., Iglesias-Fernández, R., Carbonero, P. & Matilla, Á. J. Softening-up mannan-rich cell walls. *J. Exp. Bot.* **63**, 3976–3988 (2012).
 45. Beňová-Kákošová, A. *et al.* Galactoglucomannans increase cell population density and alter the protoxylem/metaxylem tracheary element ratio in xylogenetic cultures of zinnia. *Plant Physiol.* **142**, 696–709 (2006).
 46. Moreira, L. R. S. & Filho, E. X. F. An overview of mannan structure and mannan-degrading enzyme systems. *Appl. Microbiol. Biotechnol.* **79**, 165–178 (2008).
 47. Guillotin, L., Richet, N., Lafite, P. & Daniellou, R. Is the acid/base catalytic residue mutation in β -d-mannosidase DtMan from *Dictyoglomus thermophilum* sufficient enough to provide thioglycoligase activity? *Biochimie* **137**, 190–196 (2017).
 48. Møller, M. S. Impact of Modular Architecture on Activity of Glycoside Hydrolase Family 5 Subfamily 8 Mannanases. *Molecules* vol. 27 (2022).
 49. Couturier, M. *et al.* Functional exploration of the glycoside hydrolase family GH113. *PLoS One* **17**, e0267509 (2022).
 50. Tao, F., Swarup, S. & Zhang, L. H. Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for *Xanthomonas campestris* biofilm formation. *Environ. Microbiol.* **12**, 3159–3170 (2010).
 51. Ting, T.-Y., Baharin, A., Ramzi, A. B., Ng, C.-L. & Goh, H.-H. Neprosin belongs to a new family of glutamic peptidase based on in silico evidence. *Plant Physiol. Biochem.* **183**, 23–35 (2022).
 52. Yadavalli, S. S. *et al.* Antimicrobial peptides trigger a division block in *Escherichia coli* through

- stimulation of a signalling system. *Nat. Commun.* **7**, 1–10 (2016).
53. Moreira, L. M. *et al.* Chemotactic signal transduction and phosphate metabolism as adaptive strategies during citrus canker induction by *Xanthomonas citri*. *Funct. Integr. Genomics* **15**, 197–210 (2015).
 54. Wei, C. *et al.* Global regulator PhoP is necessary for motility, biofilm formation, exoenzyme production and virulence of *Xanthomonas citri* subsp. *citri* on citrus plants. *Genes (Basel)*. **10**, 1–19 (2019).
 55. Li, J. & Wang, N. Genome-wide mutagenesis of *Xanthomonas axonopodis* pv. *citri* reveals novel genetic determinants and regulation mechanisms of biofilm formation. *PLoS One* **6**, (2011).
 56. Pauly, M. *et al.* Hemicellulose biosynthesis. *Planta* **238**, 627–642 (2013).
 57. Rennie, E. A. & Scheller, H. V. Xylan biosynthesis. *Curr. Opin. Biotechnol.* **26**, 100–107 (2014).
 58. Santos, C. R. *et al.* Molecular mechanisms associated with xylan degradation by *Xanthomonas* plant pathogens. *J. Biol. Chem.* **289**, 32186–32200 (2014).
 59. Chow, V. *et al.* Xylan utilization regulon in *Xanthomonas citri* pv. *citri* strain 306: Gene expression and utilization of oligoxylosides. *Appl. Environ. Microbiol.* **81**, 2163–2172 (2015).
 60. Solé, M. *et al.* *Xanthomonas campestris* pv. *vesicatoria* secretes proteases and xylanases via the Xps type II secretion system and outer membrane vesicles. *J. Bacteriol.* **197**, 2879–2893 (2015).
 61. Rajeshwari, R., Jha, G. & Sonti, R. V. Role of an in planta-expressed xylanase of *Xanthomonas oryzae* pv. *oryzae* in promoting virulence on rice. *Mol. Plant-Microbe Interact.* **18**, 830–837 (2005).
 62. Morais, M. A. B. *et al.* Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations. *Nat. Commun.* **12**, (2021).
 63. Luis, A. S. *et al.* Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* **3**, 210–219 (2018).
 64. Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
 65. Rogowski, A. *et al.* Evidence That GH115 α -glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. *J. Biol. Chem.* **289**, 53–64 (2014).
 66. Chuanchuen, R., Narasaki, C. T. & Schweizer, H. P. The MexJK efflux pump of *Pseudomonas aeruginosa* requires OprM for antibiotic efflux but not for efflux of triclosan. *J. Bacteriol.* **184**, 5036–5044 (2002).
 67. Chuanchuen, R., Gaynor, J. B., Karkhoff-Schweizer, R. A. & Schweizer, H. P. Molecular characterization of MexL, the transcriptional repressor of the mexJK multidrug efflux operon in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **49**, 1844–1851 (2005).
 68. Colclough, A. L., Scadden, J. & Blair, J. M. A. TetR-family transcription factors in Gram-negative bacteria: Conservation, variation and implications for efflux-mediated antimicrobial resistance. *BMC Genomics* **20**, 1–12 (2019).
 69. Keegstra, K. Plant cell walls. *Plant Physiol.* **154**, 483–486 (2010).
 70. Kaur, R., Sharma, M., Ji, D., Xu, M. & Agyei, D. Structural Features, Modification, and Functionalities of Beta-Glucan. *Fibers* vol. 8 (2020).
 71. Gluck-Thaler, E. *et al.* Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles. *Sci. Adv.* **6**, 1–11 (2020).

72. Gustavo, A.-M. *et al.* Expression profiling of virulence and pathogenicity genes of *Xanthomonas axonopodis* pv. *citri*. *J. Bacteriol.* **187**, 1201–1205 (2005).
73. Jalan, N. *et al.* Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. *BMC Genomics* **14**, (2013).
74. Rosseto, F. R., Manzone, L. R., de Oliveira Neto, M. & Polikarpov, I. Biophysical and biochemical studies of a major endoglucanase secreted by *Xanthomonas campestris* pv. *campestris*. *Enzyme Microb. Technol.* **91**, 1–7 (2016).
75. de Melo, R. R. *et al.* Identification of a cold-adapted and metal-stimulated β -1,4-glucanase with potential use in the extraction of bioactive compounds from plants. *Int. J. Biol. Macromol.* **166**, 190–199 (2021).
76. Karthik, S. *et al.* Bacterial glycogen provides short-term benefits in changing environments. *Appl. Environ. Microbiol.* **86**, e00049-20 (2020).
77. Ferreira, R. M. *et al.* Unravelling potential virulence factor candidates in *Xanthomonas citri* subsp. *citri* by secretome analysis. *PeerJ* **2016**, 1–29 (2016).
78. Soares, M. R. *et al.* Proteome of the phytopathogen *Xanthomonas citri* subsp. *citri*: A global expression profile. *Proteome Sci.* **8**, 1–11 (2010).
79. Foley, M. H. *et al.* A Cell-Surface GH9 endo-glucanase coordinates with surface glycan-binding proteins to mediate xyloglucan uptake in the gut symbiont *Bacteroides ovatus*. *J. Mol. Biol.* **431**, 981–995 (2019).
80. Pereira, C. T., Moutran, A., Fessel, M. & Balan, A. The sulfur/sulfonates transport systems in *Xanthomonas citri* pv. *citri*. *BMC Genomics* **16**, 1–11 (2015).
81. Dunger, G., Llontop, E., Guzzo, C. R. & Farah, C. S. The *Xanthomonas* type IV pilus. *Curr. Opin. Microbiol.* **30**, 88–97 (2016).
82. Nunn, D. Bacterial Type II protein export and pilus biogenesis: More than just homologies? *Trends Cell Biol.* **9**, 402–408 (1999).
83. Kogenaru, S., Qing, Y., Guo, Y. & Wang, N. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics* **13**, 1 (2012).
84. Guo, Y., Figueiredo, F., Jones, J. & Wang, N. HrpG and HrpX play global roles in coordinating different virulence traits of *Xanthomonas axonopodis* pv. *citri*. *Mol. Plant-Microbe Interact.* **24**, 649–661 (2011).
85. Poncet, S. *et al.* Correlations between carbon metabolism and virulence in bacteria. in *Contributions to Microbiology* vol. 16 88–102 (2009).
86. Facincani, A. P. *et al.* Comparative proteomic analysis reveals that T3SS, Tfp, and xanthan gum are key factors in initial stages of *Citrus sinensis* infection by *Xanthomonas citri* subsp. *citri*. *Funct. Integr. Genomics* **14**, 205–217 (2014).
87. Roeschlin, R. A. *et al.* Resistance to citrus canker induced by a variant of *Xanthomonas citri* ssp. *citri* is associated with a hypersensitive cell death response involving autophagy-associated vacuolar processes. *Mol. Plant Pathol.* **18**, 1267–1281 (2017).
88. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
89. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

90. Da Silva, A. C. R. *et al.* Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463 (2002).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
92. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, (2019).
93. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
94. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

Supplementary Information

Transcriptional landscape of *Xanthomonas citri* in response to plant carbohydrates: depicting polysaccharide utilization systems and pathogenicity modulation

Isabela Mendes Bonfim^{1,2}, Douglas Alvarez Paixão¹, José Alberto Diogo^{1,2}, Joaquim Martins Junior¹, Gabriela Felix Persinoti¹, Priscila Oliveira de Giuseppe^{1,*} and Mário Tyago Murakami^{1,*}

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPem), Zip Code 13083-970, Campinas, Sao Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

*Correspondence to priscila.giuseppe@lnbr.cnpem.br and mario.murakami@lnbr.cnpem.br

Supplementary Table 1. Summary of RNA-seq data of *X. citri* growth in minimal medium XVM2m containing different carbohydrate sources or in LBON medium. Each sample represents a biologically independent experiment.

Sample	Input reads	QC reads	QC reads (%)	rRNA reads (%)	Mapped reads (%)
XVM2m_Glucose 1	2895851	2524189	87.17%	5.64%	87.68%
XVM2m_Glucose 2	5660364	4990317	88.16%	1.92%	91.82%
XVM2m_Glucose 3	5540215	5018731	90.59%	2.58%	68.56%
XVM2m_Glucose 4	5164471	4599580	89.06%	0.79%	69.50%
XVM2m_Glucose 5	4847539	4237597	87.42%	14.36%	61.53%
XVM2m_Glucose 6	23704254	19279704	81.33%	22.76%	73.24%
XVM2m_Cellobiose 1	5169619	4564215	88.29%	6.83%	86.44%
XVM2m_Cellobiose 2	3125511	2787587	89.19%	30.06%	88.34%
XVM2m_Cellobiose 3	3742334	3170705	84.73%	0.91%	87.39%
XVM2m_Cellobiose 4	6057031	5509092	90.95%	33.62%	84.61%
XVM2m_Starch 1	8486939	7186963	84.68%	4.05%	84.98%
XVM2m_Starch 2	4554717	3935401	86.40%	3.17%	95.30%
XVM2m_Starch 3	4102919	3392366	82.68%	18.30%	94.20%
XVM2m_Starch 4	5073033	4483765	88.38%	22.24%	82.07%
XVM2m_GMOs 1	2991429	2655924	88.78%	0.69%	89.51%
XVM2m_GMOs 2	5850773	5308933	90.74%	0.79%	89.04%
XVM2m_AXOs 1	12923348	11758809	90.99%	1.47%	55.35%
XVM2m_AXOs 2	4243708	3742422	88.19%	3.78%	54.02%
XVM2m_AXOs 3	2798804	2507050	89.58%	21.03%	41.37%
XVM2m_XOs 1	8002134	7091370	88.62%	9.37%	86.27%
XVM2m_XOs 2	2786541	2382136	85.49%	64.84%	94.39%
XVM2m_XOs 3	5331341	4645216	87.13%	27.91%	79.21%
LBON 1	7182071	6247424	86.99%	6.86%	99.38%
LBON 2	5126895	4465913	87.11%	8%	99.56%
LBON 3	5641467	4955965	87.85%	1.21%	99.57%
LBON 4	16808541	14559558	86.62%	69.77%	98.53%
Medium input reads:	167811849	Medium mapped reads (%):		82.42%	

Supplementary Table 2. List of *X. citri* 306 genes assigned to glycoside hydrolase (GH) families discussed in this study. ^a Biochemical function characterized in *X. citri* or *Xanthomonas* sp. orthologs as reported in the references. ^b Putative function based on the description available on KEGG¹ or Uniprot databases².

ID	Description	CAZy	Biochemical function ^{a,b}	Reference
XAC3312	Glycosyl hydrolase	GH2	Not assigned	-
XAC4250	β -Galactosidase	GH2	Hydrolysis of terminal non-reducing β -D-galactose residues in β -D-galactosides ^b	-
XAC1334	N-acetyl- β -glucosaminidase	GH3	Hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl- β -D-hexosaminides ^b	-
XAC1448	β - Glucosidase	GH3	Exo-hydrolysis of β -glucosidic linkages in oligos or polysaccharides with the release of glucose ^a	3
XAC1793	β - Glucosidase	GH3	Exo-hydrolysis of β -glucosidic linkages in oligos or polysaccharides with the release of glucose ^a	3
XAC3076	β - Xylosidase	GH3	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^a	3
XAC3869	β - Glucosidase	GH3	Exo-hydrolysis of β -glucosidic linkages in oligos or polysaccharides with the release of glucose ^a	3
XAC4231	β -Xylosidase	GH3	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^a	3
XAC3081	6-phospho- β -glucosidase	GH4	Catalyze the hydrolysis of 6-phospho- β -D-glucosyl-(1 \rightarrow 4)-D-glucose in D-glucose and D-glucose-6-phosphate ^b	-
XAC0028	Cellulase	GH5_5	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose, lichenin or cereal β -glucans ^b	-
XAC0029	Cellulase	GH5_5	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose, lichenin or cereal β -glucans ^b	-
XAC0030	Cellulase	GH5_5	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose, lichenin or cereal β -glucans ^b	-

Supplementary Table 2. Continued.

ID	Description	CAZy	Biochemical function ^{a,b}	Reference
XAC0346	Degenerated cellulase	GH5	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose, lichenin or cereal β -glucans ^b	-
XAC0612	Cellulase	GH5_1	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose ^a	4
XAC1796	Endo- β -1,4-mannanase	GH5_8	Endo-hydrolysis of β -1,4-D-mannosidic linkages in galactomannans and glucomannans; biofilm dispersal ^a	5,6,7
XAC3516	Cellulase	GH8	Endo-hydrolysis of β -1,4-D-glucosidic linkages in cellulose ^a	8
XAC2522	Cellulase	GH9	Endo-hydrolysis of xyloglucan ^a ; is also predicted to cleave β -1,4-D-glucosidic linkages in cellulose, lichenin or cereal β -glucans ^b	3
XAC4249	Exo-oligoxyylanase	GH10	Hydrolysis of xylo-oligosaccharides with the release of xylose ^a	9
XAC4252	Xylanase	GH10	Endo-hydrolysis of β -1,4-D-xylosidic linkages in xylans ^b	-
XAC4254	Endo- β -1,4-xylanase	GH10	Endo-hydrolysis of β -1,4-D-xylosidic linkages in xylans ^a	9
XAC0154	Maltosyltransferase	GH13_3	Uses maltose-1-phosphate as the sugar donor to elongate linear or branched α -(1-4)-glucans ^b	-
XAC0155	Trehalose synthase	GH13_16	Generates trehalose from maltose ^b	-
XAC0156	1,4- α -Glucan branching enzyme	CBM48, GH13_9	Catalyzes the formation of the α -1,6-glucosidic linkages in glycogen by scission of a 1,4-alpha-linked oligosaccharide from growing glucan chains ^b	-
XAC0426	1,4- α -Glucan branching enzyme	CBM48, GH13_9	Catalyzes the formation of the α -1,6-glucosidic linkages in glycogen by scission of a 1,4-alpha-linked oligosaccharide from growing glucan chains ^b	-

Supplementary Table 2. Continued.

ID	Description	CAZy	Biochemical function ^{a,b}	Reference
XAC0427	Malto-oligosyltrehalose trehalohydrolase	CBM48, GH13_10	Hydrolysis of α -(1-4)-D-glycosidic linkage in α -4-D-((α -(1-4)-glucanosyl) (n) trehalose to yield trehalose and α -(1-4)-D -glucan ^b	-
XAC0429	Maltooligosyltrehalose synthase	GH13_26	Involved in trehalose biosynthesis, converting the α -1,4-glucosidic linkage to an α , α -1,1-glucosidic linkage at the reducing end of the maltooligosaccharide ^b	-
XAC0431	Glycogen debranching enzyme	CBM48, GH13_11	Debranching enzyme for glycogen catabolism ^b	-
XAC0798	α -Amylase	GH13_27	Endohydrolysis of α -(1 \rightarrow 4)-D-glucosidic linkages in polysaccharides containing three or more α -(1 \rightarrow 4)-linked D-glucose units ^a	10
XAC2596	Cyclomaltodextrin glucanotransferase	GH13	Catalyzes the reaction of cyclizing part of a 1,4- α -D-glucan molecule, generating non-reducing cyclic dextrans or cyclodextrans ^b	-
XAC2602	α -Glucosidase/ α -glucosyl transferase	GH13_23	Hydrolysis of glucosidic linkages in maltose or α -glycosylation activity toward alcoholic and phenolic -OH groups, using maltose as an α -glucosyl donor ^a	11,12
XAC3254	Glycogen debranching enzyme	CBM48, GH13_11	Debranching enzyme for glycogen catabolism ^b	-
XAC3490	Sucrose hydrolase	GH13_4	Hydrolysis of sucrose ^b	-
XAC1177	Hypothetical protein	GH15	Not assigned	-
XAC3210	Hypothetical protein	GH15	Not assigned	-
XAC4082	Hypothetical protein	GH15	Not assigned	-
XAC1176	Glycosyl hydrolase	GH30	Not assigned	-

Supplementary Table 2. Continued.

ID	Description	CAZy	Biochemical function ^{a,b}	Reference
XAC0165	Xylosidase	GH43	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^b	-
XAC1275	α -L-arabinofuranosidase	GH43	Hydrolysis of non-reducing α -L-arabinofuranoside residues in α -L-arabinosides ^b	-
XAC2533	Xylosidase	GH43	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^b	-
XAC4183	α -L-arabinofuranosidase	GH43	Hydrolysis of non-reducing α -L-arabinofuranoside residues in α -L-arabinosides ^b	-
XAC4230	Xylosidase	GH43	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^b	-
XAC4258	Exo-oligoxylanase	GH43	Hydrolysis of xylo-oligosaccharides and xylan with the release of xylose ^a	13
XAC4227	α -Glucuronidase	GH67	Hydrolysis of the α -1,2-glycosidic bond between 4-O-methyl α -glucuronic acid and the terminal nonreducing end xylopyranosyl unit of small xylo-oligosaccharides ^b	-
XAC2599	α -Glucosidase	GH97	Exo-hydrolysis of α -glucosidic linkages in oligos or polysaccharides with the release of glucose ^b	-
XAC3313	α -Glucosidase	GH97	Exo-hydrolysis of α -glucosidic linkages in oligos or polysaccharides with the release of glucose ^b	-
XAC3522	Endo- β -1,4-mannanase	GH113	Random hydrolysis of β -1,4-D-mannosidic linkages in mannans ^a	14
XAC4194	conserved hypothetical protein	GH115	Not assigned	-

Supplementary Table 3. Sequence-based prediction of subcellular localization of proteins putative related to starch depolymerization in *X. citri*. The position of the start codon originally annotated for these ORFs was revised based on the reference sequences whose accession codes are shown in NR ID. TMH = transmembrane helix. CYT = cytoplasm, EXT = extracellular, PER = periplasm, OM = outer membrane.

Enzymes encoded by the Starch PUL								
<i>Locus</i>	NR ID	Description	SignalP 6.0 ¹⁵	TMHMM ¹⁶	CELLO v.2.5 ¹⁷	SOSUI/GramN ¹⁸	Consensus	Criterion
XAC0798	WP_015462876.1	α -Amylase	Sec/SPI (1-14)	-	EXT	EXT	EXT	SPI site accessible
XAC2596	WP_011051695.1	Cyclomaltodextrin glucanotransferase	Other	-	PER	CYT	CYT	Lack of signal peptide
XAC2598	WP_011051697.1	Conserved hypothetical protein	Sec/SPI (1-21)	-	OM	PER	PER	SPI site accessible/PER
XAC2599	WP_011051698.1	α -Glucosidase	Sec/SPI (1-35)	-	CYT	PER	PER	SPI site accessible/PER
XAC2602	WP_011051701.1	α -Glucosidase	-	-	CYT	CYT	CYT	Lack of signal peptide

Supplementary Table 4. Over Representation Analysis (ORA) of biochemical pathways for the starch condition using glucose as reference.

ID	Description	Gene Ratio	Bg Ratio	p-value	p-adjust	q-value	Gene ID	Count
xac02030	Bacterial chemotaxis	27/70	48/1340	1.08e-24	3.24e-23	2.62e-23	XAC1893/XAC1902/XAC1666/XAC1908/ XAC1891/XAC1746/XAC2447/XAC1899/ XAC0611/XAC1906/XAC1930/XAC1897/ XAC1896/XAC1903/XAC1996/XAC2448/ XAC1932/XAC1900/XAC1931/XAC1888/ XAC1904/XAC1892/XAC1987/XAC3132/ XAC1894/XAC1909/XAC1890	27
xac02020	Two-component system	33/70	146/1340	1.88e-15	2.82e-14	2.28e-14	XAC0346/XAC2982/XAC2983/XAC1893/ XAC1577/XAC4023/XAC1902/XAC1666/ XAC1891/XAC1746/XAC2447/XAC1899/ XAC0611/XAC1906/XAC1930/XAC1897/ XAC1896/XAC1903/XAC1996/XAC2448/ XAC1932/XAC1900/XAC1888/XAC1933/ XAC1904/XAC1892/XAC1987/XAC1975/ XAC3132/XAC1894/XAC1909/XAC1989/ XAC1890	33
xac00920	Sulfur metabolism	12/70	31/1340	1.03e-08	1.03e-07	8.30e-08	XAC3040/XAC0830/XAC3341/XAC3039/ XAC0827/XAC0849/XAC0848/XAC0829/ XAC0828/XAC3332/XAC3331/XAC3330	12
xac02040	Flagellar assembly	8/70	41/1340	0.0009	0.007	0.005	XAC1908/XAC1988/XAC1974/XAC1933/ XAC1973/XAC1975/XAC1909/XAC1989	8

Supplementary Table 5. Hits found in BLAST searches against the Protein Data Bank database using the sequences of enzymes probably related to galactomannan depolymerization in *X. citri*.

Bait protein	BLAST HIT (PDB code)	Biochemical function	Organism	Sequence ID	E-value	Query cover	Reference
XAC1796	4FK9	Catalytic domain of β -mannanase	<i>Streptomyces sp. Sirex</i> AA-E	49.3%	8.0E-91	90%	¹⁹
XAC3312	5N6U	β -D-Mannosidase	<i>Dictyoglomus thermophilum</i> H-6-12	24.2%	1.0E-42	81%	²⁰
XAC3313	3A24	α -Galactosidase	<i>Bacteroides thetaiotaomicron</i>	33.6%	1.0E-112	89%	²¹
XAC3314	-	-	-	-	-	-	-
XAC3315	1K4Y	Carboxylesterase	<i>Oryctolagus cuniculus</i>	33.5%	9.0E-68	93%	²²

Supplementary Table 6. Sequence-based prediction of subcellular localization of proteins putative related to mannan depolymerization in *X. citri*. The position of the start codon originally annotated for these ORFs was revised based on the reference sequences whose accession codes are shown in NR ID. TMH = transmembrane helix. CYT = cytoplasm, EXT = extracellular, PER = periplasm, OM = outer membrane, IM = inner membrane. ^a SPII signal predicts for lipoproteins, i.e membrane-tethered proteins via N-terminal lipidation. ^b Position +2 different from D (aspartic acid) after a SPII cleavage site indicates outer membrane anchoring²³.

Enzymes encoded by the Mannan PUL								
<i>Locus</i>	NR ID	Description	SignalP 6.0 ¹⁵	TMHMM ¹⁶	CELLO v.2.5 ¹⁷	SOSUI/GramN ¹⁸	Consensus	Criterion
XAC1796	WP_003481985.1	Endo- β -1,4-mannanase	Sec/SPI (1-26)	-	EXT	EXT	EXT	SPI site accessible
XAC3312	WP_015471689.1	β -D-Mannosidase	Sec/SPI (1-27)	-	OM	-	-	Inconclusive
XAC3313	WP_011052186.1	α -Galactosidase	Sec/SPI (1-25)	-	OM	PER	PER	SPI site accessible/PER
XAC3314	WP_005929971.1	Hypothetical protein	-	-	EXT	CYT	CYT	Lack of signal peptide
XAC3315	WP_011052187.1	Carboxylesterase	Sec/SPI (1-21)	-	IM	PER	PER	SPI site accessible/PER
XAC3522	WP_015472685.1	Hypothetical protein	Sec/SPII (1-20) +2=N	-	PER	PER	OM	N-terminal lipid anchor/ +2 \neq D ^b

Supplementary Table 7. Sequence-based prediction of subcellular localization of proteins putative related to xylan depolymerization in *X. citri*. The position of the start codon originally annotated for these ORFs was revised based on the reference sequences whose accession codes are shown in NR ID. TMH = transmembrane helix. CYT = cytoplasm, EXT = extracellular, PER = periplasm, OM = outer membrane, IM = inner membrane.

Enzymes encoded by the Xylan PUL								
<i>Locus</i>	NR ID	Description	SignalP 6.0 ¹⁵	TMHMM ¹⁶	CELLO v.2.5 ¹⁷	SOSUI/GramN ¹⁸	Consensus	Criterion
XAC4227	WP_015463770.1	α -Glucuronidase	Sec/SPI (1-34)	1 TMH (7-29)	PER	PER	PER	SPI site accessible/PER
XAC4228	WP_011052823.1	Acetyl esterase	Sec/SPI (1-24)	-	OM	OM	OM	SPI site accessible
XAC4230	WP_040107820.1	β - Xylosidase	Sec/SPI (1-23)	-	EXT	EXT	EXT	SPI site accessible
XAC4231	WP_011052826.1	β - Xylosidase	Sec/SPI (1-29)	-	PER	PER	PER	SPI site accessible/PER
XAC4249	WP_040107704.1	Exo-oligoxyranase	Sec/SPI (1-29)	1 TMH (12-30)	PER	PER	PER	SPI site accessible/PER
XAC4250	WP_011052842.1	β -Glucuronidase	-	-	PER	CYT	CYT	No signal peptide
XAC4252	WP_003485050.1	Xylanase	Sec/SPI (1-22)	1 TMH (7-29)	PER	PER	PER	SPI site accessible/PER
XAC4254	WP_003485050.1	Xylanase	Sec/SPI (1-22)	1 TMH (7-29)	PER	PER	PER	SPI site accessible/PER
XAC4258	WP_01102848.1	β - Xylosidase	-	-	CYT	CYT	CYT	No signal peptide
XAC4194	WP_040107701.1	α -Glucuronidase	Sec/SPI (1-35)	1 TMH (20-42)	PER	CYT	PER	SPI site accessible/PER

Supplementary Table 8. List of XAC4194 orthologs in other *Xanthomonas* species according to KEGG database¹.

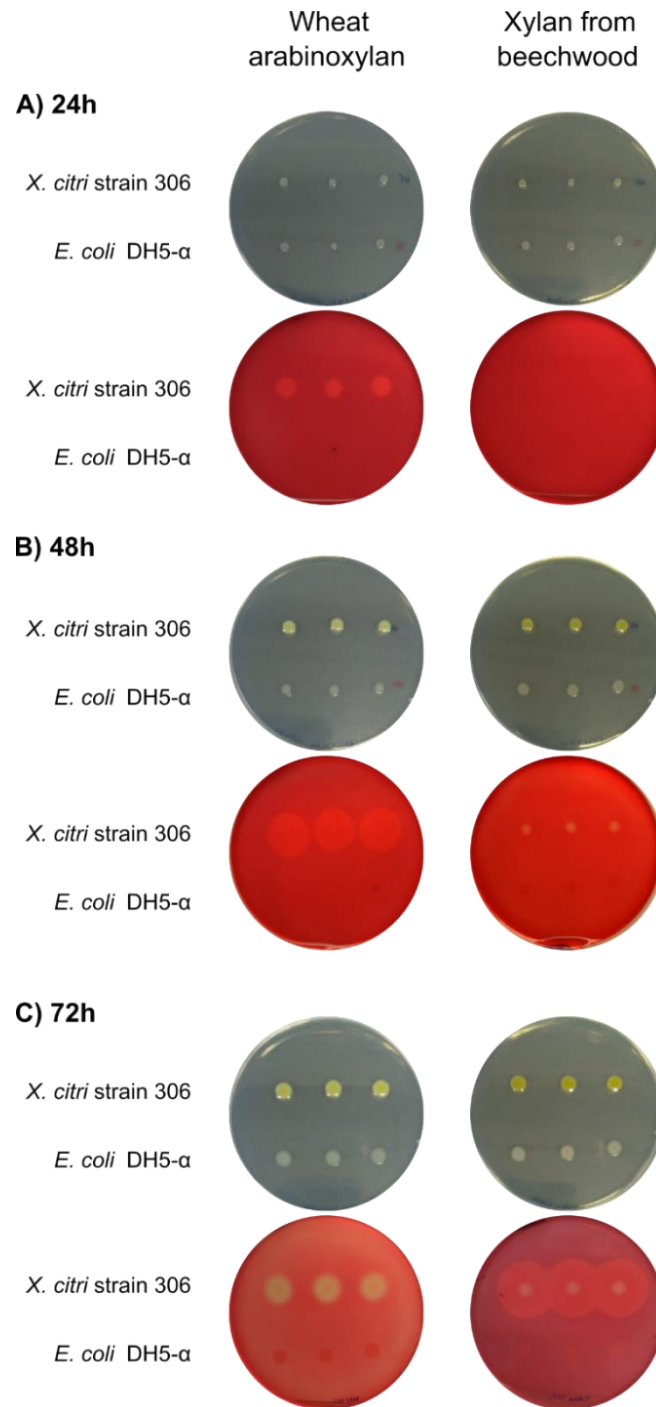
Organism	Gene ID	Identity
<i>Xanthomonas citri</i> subsp. <i>citri</i> Aw12879	XCAW_00100	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> mf20	J172_04367	100%
<i>Xanthomonas citri</i> subsp. <i>citri</i> UI6	J158_04330	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> MN12	J164_04327	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> NT17	J169_04374	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> MN11	J163_04327	100%
<i>Xanthomonas citri</i> subsp. <i>citri</i> A306	J151_04377	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> UI7	J159_04326	100%
<i>Xanthomonas citri</i> pv. <i>citri</i> MN10	J162_04332	100%
<i>Xanthomonas axonopodis</i> Xac29-1	XAC29_21140	100%
<i>Xanthomonas citri</i> pv. <i>fuscans</i>	XFF4834R_chr40650	97.20%
<i>Xanthomonas axonopodis</i> pv. <i>citrumelo</i> F1	XACM_4067	96.60%
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i>	XCV4294	96.10%
<i>Xanthomonas</i> sp. ISO98C4	AC801_20900	95.20%
<i>Xanthomonas phaseoli</i>	XppCFBP6546_13385	95.20%
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC 10331	XOO0335	92.70%
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A	PXO_02750	93.30%
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP7342	BE73_22620	93.50%
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	XOC_0357	93.50%
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	XOO0307	93.40%
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO86	AZ54_23445	93.30%
<i>Xanthomonas vasicola</i> pv. <i>vasculorum</i>	C7V42_21405	90%

Supplementary Table 9. Sequence-based prediction of subcellular localization of proteins putative related to β -glucans depolymerization in *X. citri*. The position of the start codon originally annotated for these ORFs was revised based on the reference sequences whose accession codes are shown in NR ID. TMH = transmembrane helix. CYT = cytoplasm, EXT = extracellular, PER = periplasm, OM = outer membrane, IM = inner membrane. ^a SPII signal predicts for lipoproteins, i.e membrane-tethered proteins via N-terminal lipidation. ^b Position +2 different from D (aspartic acid) after a SPII cleavage site indicates outer membrane anchoring²³.

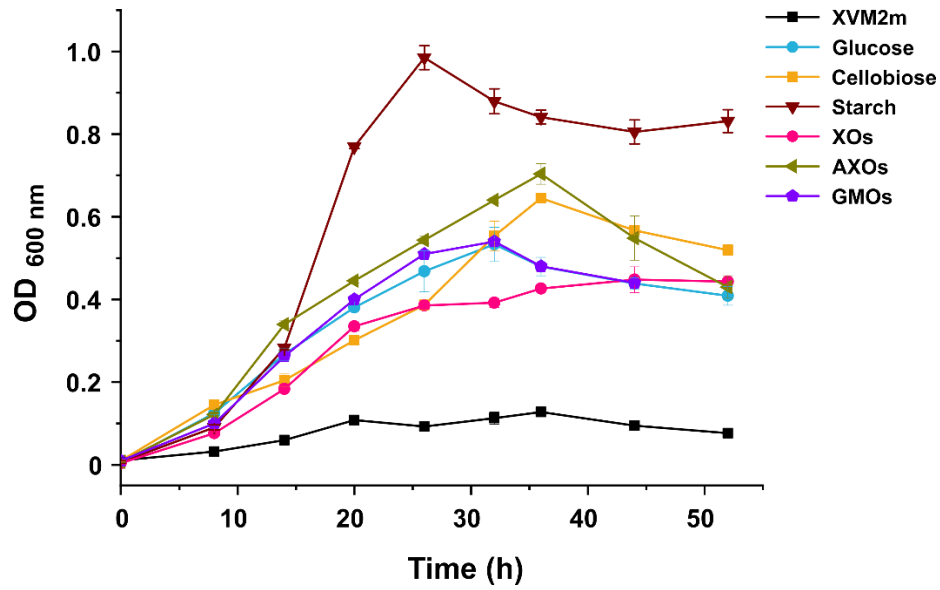
Enzymes encoded by the cellulose and mixed-linked β -glucan utilization loci								
<i>Locus</i>	NR ID	Description	SignalP 6.0 ¹⁵	TMHMM ¹⁶	CELLO v.2.5 ¹⁷	SOSUI/GramN ¹⁸	Consensus	Criterion
XAC0028	WP_040280431.1	Endo- β -1,4-glucanase	Sec/SPI (1-29)	-	OM	OM	EXT	SPI site accessible and literature ²⁴
XAC0029	WP_003482146.1	Endo- β -1,4-glucanase	Sec/SPI (1-29)	-	OM	EXT	EXT	SPI site accessible and literature ²⁴
XAC0030	WP_011050034.1	Endo- β -1,4-glucanase	Sec/SPI (1-29)	-	PER	PER	PER	SPI site accessible and literature ²⁵
XAC0346	WP_078562049.1	Degenerated cellulase	Tat/SPI (1-52)	-	PER	PER	PER	SPI site accessible/PER
XAC0612	WP_078562008.1	Endo- β -1,4-glucanase	Sec/SPI (1-25)	-	EXT	EXT	EXT	SPI site accessible and literature ²⁴
XAC2522	WP_011051647.1	Endo- β -1,4-glucanase	Sec/SPII ^a (1-19)/ +2=A	-	PER	PER	OM	N-terminal lipid anchor/ +2 \neq D ^b
XAC3869	WP_011052578.1	β -Glucosidase	-	-	CYT	CYT	CYT	Lack of signal peptide

Supplementary Table 10. Conditions for enzymatic hydrolysis of different polysaccharides in XVM2m minimal medium using recombinant proteins of *X. citri*. All the substrates were provided by Megazyme.

Substrate	Enzyme	Final concentration	pH	T °C	Time
Xylan (<i>from beechwood</i>)	XAC4254	0.03 mg. mL ⁻¹	6	30 °C	14 h
Arabinoxylan (<i>wheat flour</i>)	XAC4254	0.004 mg. mL ⁻¹	6	30 °C	14 h
Galactomannan (<i>carob</i>)	XAC3522	0.3 mg. mL ⁻¹	5.5	35 °C	5 h



Supplementary Figure 1. Test of polysaccharides degradation by colony growing on NYG-agar plates with wheat arabinoxylan or beechwood xylan. Three colonies of *X. citri* 306 (first line) and three of *E. coli* DH5 α (second line) were inoculated on different plates. After 24 (A), 48(B) and 72(C) hours a plate of each carbohydrate was photographed, colonies removed and then stained with congo red. Plates with wheat arabinoxylan are in the left column and those with beechwood xylan are in the right. The clear regions on the stained plates indicate the depolymerization of the polysaccharides by *X. citri*. *E. coli* DH5 α was used as a negative control.



Supplementary Figure 2. Growth curves analysis of *X. citri* in minimal medium XVM2m with different carbohydrates sources. Data are presented as mean \pm SD from three independent experiments (n=3). XOs – xylan oligosaccharides; AXOs – arabinoxylan oligosaccharides; GMOs – galactomannan oligosaccharides.

References

1. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
2. Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
3. Vieira, P. S. *et al.* Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* **12**, 1–15 (2021).
4. Rosseto, F. R., Manzine, L. R., de Oliveira Neto, M. & Polikarpov, I. Biophysical and biochemical studies of a major endoglucanase secreted by *Xanthomonas campestris* pv. *campestris*. *Enzyme Microb. Technol.* **91**, 1–7 (2016).
5. Dow, J. M. *et al.* Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10995–11000 (2003).
6. Hsiao, Y. M., Liu, Y. F., Fang, M. C. & Tseng, Y. H. Transcriptional regulation and molecular characterization of the *manA* gene encoding the biofilm dispersing enzyme mannan endo-1, 4- β -mannosidase in *Xanthomonas campestris*. *J. Agric. Food Chem.* **58**, 1653–1663 (2010).
7. Møller, M. S. Impact of modular architecture on activity of glycoside hydrolase family 5 subfamily 8 mannanases. *Molecules* vol. 27 (2022).
8. de Melo, R. R. *et al.* Identification of a cold-adapted and metal-stimulated β -1,4-glucanase with potential use in the extraction of bioactive compounds from plants. *Int. J. Biol. Macromol.* **166**, 190–199 (2021).
9. Santos, C. R. *et al.* Molecular mechanisms associated with xylan degradation by *Xanthomonas* plant pathogens. *J. Biol. Chem.* **289**, 32186–32200 (2014).
10. Lin, Y. *et al.* Extracellular amylase is required for full virulence and regulated by the global posttranscriptional regulator *rsma* in *Xanthomonas campestris* pathovar *campestris*. *Phytopathology* **111**, 1104–1113 (2021).
11. Sato, T. *et al.* Purification, characterization, and gene identification of an α -glucosyl transfer enzyme, a novel type α -glucosidase from *Xanthomonas campestris* WU-9701. *J. Mol. Catal. B Enzym.* **80**, 20–27 (2012).
12. Watanabe, R., Arimura, Y., Ishii, Y. & Kirimura, K. Crystal structure of α -glucosyl transfer enzyme XgtA from *Xanthomonas campestris* WU-9701. *Biochem. Biophys. Res. Commun.* **526**, 580–585 (2020).
13. Morais, M. A. B. *et al.* Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations. *Nat. Commun.* **12**, (2021).
14. Couturier, M. *et al.* Functional exploration of the glycoside hydrolase family GH113. *PLoS One* **17**, e0267509 (2022).
15. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, (2022).
16. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
17. Yu, C.-S., Chen, Y.-C., Lu, C.-H. & Hwang, J.-K. Prediction of protein subcellular localization. *Proteins Struct. Funct. Bioinforma.* **64**, 643–651 (2006).

18. Imai, K. *et al.* SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. *Bioinformatics* **2**, 417–421 (2008).
19. Takasuka, T. E. *et al.* Biochemical properties and atomic resolution structure of a proteolytically processed β -mannanase from cellulolytic *Streptomyces* sp. SirexAA-E. *PLoS One* **9**, e94166 (2014).
20. Guillotin, L., Richet, N., Lafite, P. & Daniellou, R. Is the acid/base catalytic residue mutation in β -d-mannosidase *DtMan* from *Dictyoglomus thermophilum* sufficient enough to provide thioglycoligase activity? *Biochimie* **137**, 190–196 (2017).
21. Okuyama, M. *et al.* Catalytic mechanism of retaining α -galactosidase belonging to glycoside hydrolase family 97. *J. Mol. Biol.* **392**, 1232–1241 (2009).
22. Bencharit, S. *et al.* Structural insights into CPT-11 activation by mammalian carboxylesterases. *Nat. Struct. Biol.* **9**, 337–342 (2002).
23. Seydel, A., Gounon, P. & Pugsley, A. P. Testing the ‘+2 rule’ for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol. Microbiol.* **34**, 810–821 (1999).
24. Ferreira, R. M. *et al.* Unravelling potential virulence factor candidates in *Xanthomonas citri* subsp. *citri* by secretome analysis. *PeerJ* **2016**, 1–29 (2016).
25. Soares, M. R. *et al.* Proteome of the phytopathogen *Xanthomonas citri* subsp. *citri*: A global expression profile. *Proteome Sci.* **8**, 1–11 (2010).

Chapter 3

Type of chapter: Research Article

Journal (Publisher): Nature Communications (Springer Nature)

Impact factor: 17.69 (2021)

Date of publication: June 30, 2021

DOI: 10.1038/s41467-021-24277-4

Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors

Plinio S. Vieira^{1#}, Isabela M. Bonfim^{1,2#}, Evandro A. Araujo^{1,3}, Ricardo R. Melo¹, Augusto R. Lima¹, Melissa R. Fessel⁴, Douglas A. A. Paixão¹, Gabriela F. Persinoti¹, Silvana A. Rocco⁵, Tatiani B. Lima¹, Renan A. S. Pirolla¹, Mariana A. B. Morais¹, Jessica B. L. Correa¹, Leticia M. Zanthorlin¹, Jose A. Diogo^{1,2}, Evandro A. Lima¹, Adriana Grandis⁶, Marcos S. Buckeridge⁶, Fabio C. Gozzo⁷, Celso E. Benedetti⁵, Igor Polikarpov⁸, Priscila O. Giuseppe^{1*} and Mario T. Murakami^{1*}

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil.

³Brazilian Synchrotron Light Laboratory (LNLS), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

⁴Butantan Institute, Butantan Foundation, São Paulo, São Paulo, Brazil.

⁵Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

⁶Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

⁷Institute of Chemistry, University of Campinas, Campinas, São Paulo, Brazil.

⁸São Carlos Institute of Physics, University of São Paulo, São Carlos, São Paulo, Brazil.

#These authors have equally contributed to this work

*Correspondence and requests should be addressed to M.T.M (mario.murakami@lnbr.cnpem.br) or P.O.G (priscila.giuseppe@lnbr.cnpem.br)

Abstract

Xyloglucans are highly substituted and recalcitrant polysaccharides found in the primary cell walls of vascular plants, acting as a barrier against pathogens. Here, we reveal that the diverse and economically relevant *Xanthomonas* bacteria are endowed with a xyloglucan depolymerization machinery that is linked to pathogenesis. Using the citrus canker pathogen as a model organism, we show that this system encompasses distinctive glycoside hydrolases, a modular xyloglucan acetylsterase and specific membrane transporters, demonstrating that plant-associated bacteria employ distinct molecular strategies from commensal gut bacteria to cope with xyloglucans. Notably, the sugars released by this system elicit the expression of several key virulence factors, including the type III secretion system, a membrane-embedded apparatus to deliver effector proteins into the host cells. Together, these findings shed light on the molecular mechanisms underpinning the intricate enzymatic machinery of *Xanthomonas* to depolymerize xyloglucans and uncover a role for this system in signaling pathways driving pathogenesis.

Keywords: Xyloglucan, *Xanthomonas*, CAZymes, virulence factors, type III secretion system.

Introduction

Xyloglucans (XyGs) comprise a class of highly complex polysaccharides present in the primary cell wall of vascular plants from clubmosses to angiosperms, including all agricultural cultivars¹. These recalcitrant polysaccharides form an intricate network with cellulose, which is critical for cell wall function and structure, and serves as a physical barrier against pathogen invasion and colonization².

XyGs are structurally and chemically diverse, consisting of a β -1,4-linked glucan backbone decorated with α -1,6-xylosyl residues, which might have additional decorations such as D-galactose, L-fucose, and L-arabinose, depending on the source at the tissue level in plants³. These polysaccharides can also be acetylated and this modification is known to affect their physicochemical properties and interaction with other cell-wall components^{1,3,4}.

To cope with XyGs, many microorganisms, such as saprophytes⁵ and commensal bacteria from the human gut⁶, harbor enzymatic toolboxes encoded by a set of physically linked genes known as XyG utilization loci (XyGUL). Plant pathogens from the *Xanthomonas* genus also encompass a gene cluster predicted to degrade XyGs (Fig. 1). These pathogens exhibit a high tissue and host specificity, colonizing mesophylls or xylem vessels of over 400 distinct monocotyledons and dicotyledons, including many economically important plants such as citrus, cotton and corn^{7,8}. Regardless of the lifestyle and ecological niche specialization, most *Xanthomonas* species harbors this predicted XyGUL in their genomes, indicating the relevance of this system for these bacteria. However, the molecular mechanisms underpinning XyG depolymerization and potential biological roles in *Xanthomonas* and other phytopathogens remain so far elusive.

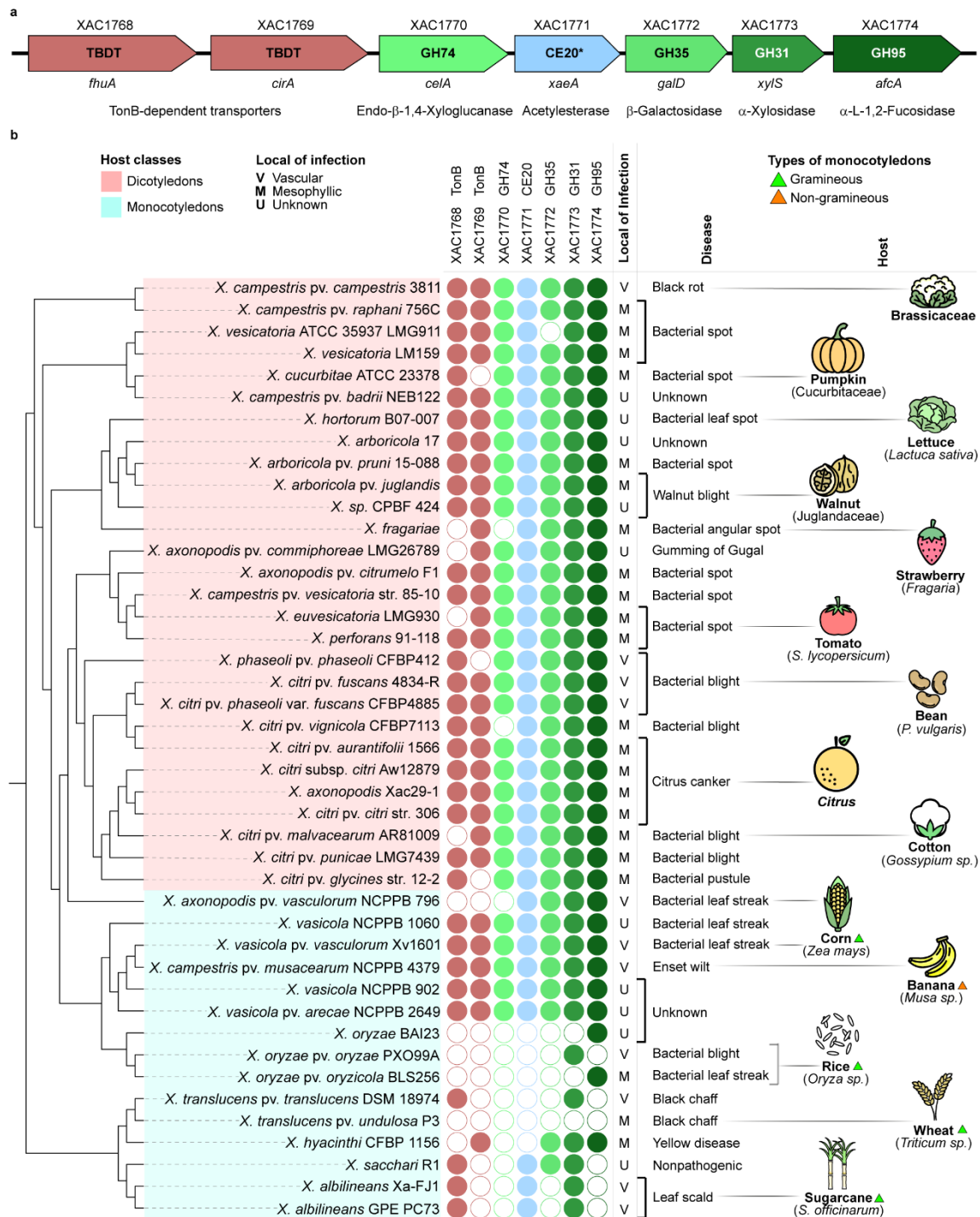


Fig. 1: The XyGUL conservation in *Xanthomonas* spp. a, XyGUL predicted in *X. citri* pv. *citri* 306 genome showing TonB-dependent transporters (TBDDT in pink), glycoside hydrolases (GHs in shades of green) and a carbohydrate esterase (CE in blue, * revealed in this work). b, Dendrogram of *Xanthomonas* species based on phylogenetic analysis (details in Supplementary Fig. 2) showing the presence (filled circles) or absence (open circles) of XyGUL genes and information about disease and tissue/host specificity. V=vascular, M=mesophyllic and U=unknown. Some species infect dicotyledons (pink box), whereas others colonize monocotyledons (blue box). Monocotyledons are subdivided in gramineous (green triangle) or non-gramineous (orange triangle).

Therefore, here we used the causal agent of citrus canker *Xanthomonas citri* pv. *citri* (*X. citri*) as a model organism⁹ to investigate the molecular basis of XyG breakdown and its potential involvement in pathogenesis and host-pathogen interactions. Our results show that *Xanthomonas* XyGUL encodes a highly elaborated enzymatic cascade including distinct activities (acetylsterase, α -L-fucosidase, β -galactosidase, α -xylosidase and xyloglucanase), catalytic mechanisms (inverting and retaining), modes of action (endo and exo) and 3D architectures (multi-modular and quaternary arrangements). This machinery notably differs from other known XyGULs, expanding the current knowledge about microbial molecular strategies associated with the depolymerization and utilization of recalcitrant plant polysaccharides. Furthermore, we reveal a link between this enzymatic system and bacterial virulence through a stimulatory effect of its products on the expression of several key virulence factors, including the type III secretion system (T3SS), a needle-like apparatus that inject effector proteins into the plant cell to modulate host responses in favor of bacterial colonization¹⁰.

Results

The XyGUL gene architecture in *Xanthomonas* diverges from Bacteroidetes

Genomic analysis revealed that most *Xanthomonas* species conserve a predicted XyGUL consisting of two outer membrane TonB-dependent transporters (TBDTs), four glycoside hydrolases belonging to the families GH74, GH31, GH35, GH95 and one esterase with no significant similarity with any carbohydrate esterase family (Fig. 1a, and Supplementary Tables 1 and 2). Adjacent to this cluster, there are common components for xylose metabolism (D-xylulokinase and xylose isomerase) and an inner membrane MFS sugar transporter (Supplementary Figs. 1 and 2). A search in the polysaccharide utilization loci database (PULDB¹¹) did not result in any similar organization in Bacteroidetes, except for the clustering of three or two XyG-related genes in some species. In characterized Bacteroides XyGULs, the clustering of GH31 and GH95 genes has been observed, but in association with other carbohydrate-active enzymes (CAZymes)^{6,12}. The synteny of GH31, GH35 and GH95 genes was reported in the XyGUL from the saprophyte *Cellvibrio japonicus*⁵, but not physically linked to endoxyloglucanases or esterases genes.

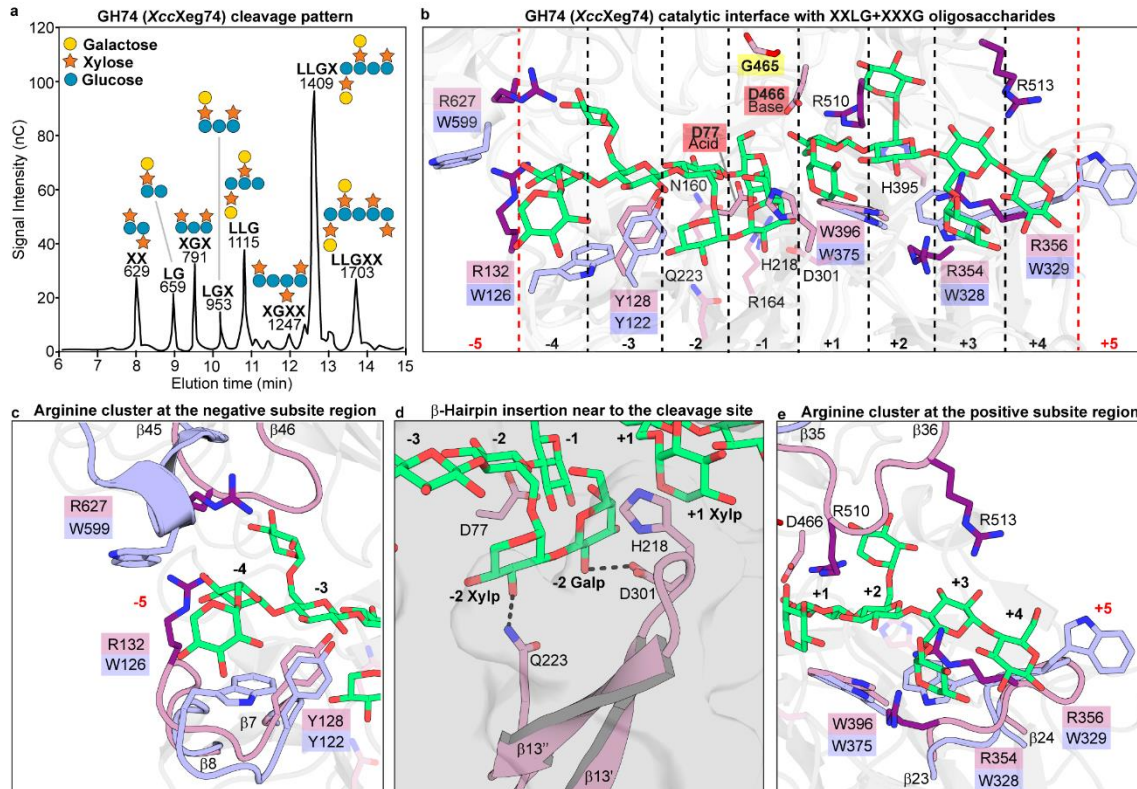


Fig. 2: The structural determinants for substrate recognition in *Xanthomonas* GH74 endo- β -1,4-xyloglucanases. a, HPAEC-PAD analysis of products released from *Copaifera langsdorffii* XyG by *XccXeg74*. Numbers above the peaks represent m/z values of each product assessed by mass spectrometry (details in Supplementary Fig. 3c). Letters indicate the type of substitutions appended to the glucose backbone: G = non-substituted glucose, X = glucose substituted with a xylose at C-6 and L = X with a galactose appended at xylose C-2. Symbols represent glucose (blue circle), xylose (orange star), and galactose (yellow circle). b, *XccXeg74* crystallographic structure (carbon atoms in dark and light purple) superimposed onto the *C. lactoaceticus* GH74 enzyme (carbon atoms in light blue; PDB ID 6P2M¹⁹), and *N. koreensis* GH74 in complex with XyGOs (only XyGOs shown; carbon atoms in green; PDB ID 6P2L¹⁹) highlighting the substrate-binding subsites (-4 to +4 in *XccXeg74*) and (-5 to +5 in *ClGH74A*). Residues involved in substrate recognition are shown as sticks, with the labels of catalytic residues from *XccXeg74* highlighted with red boxes and that of the glycine featuring acceptance to X and G motifs indicated with a yellow box. Note the arginine residues (carbon atoms in dark purple) populating the several substrate-binding subsites in *XccXeg74*. c, d, e, Amplified view of specific zones of the GH74 catalytic interface highlighting structural adaptations found in *XccXeg74* that are divergent from *ClGH74A*.

The XyGUL is highly conserved across the *Xanthomonas* genus regardless of the broad range of hosts (monocotyledons and dicotyledons) and tissue specificity (mesophyll or xylem vessels) (Fig. 1b). Few of them have lost the predicted GH74 xyloglucanase, but

endoglucanases encoded outside the XyGUL, such as GH5, GH9 and GH12 members, may compensate its absence (Supplementary Fig. 2). The only exceptions that lack most of XyGUL genes are *Xanthomonas* species colonizing gramineous monocotyledons such as *X. oryzae* (rice), *X. translucens* (wheat) and *X. albilineans* (sugarcane) (Fig. 1b). This apparent loss of XyG-degrading capacity correlates with the typically lower contents of XyG in the cell walls of these plants^{13–16}. Furthermore, many *Xanthomonas* species carrying the predicted XyGUL are promoters of several diseases in highly relevant agricultural crops such as corn (*X. vasicola* pv. *vasculorum* causing bacterial leaf streak), tomato (*X. perforans* causing bacterial spot), banana (*X. campestris* pv. *musacearum* causing enset wilt), citrus (*X. citri* pv. *citri* causing citrus canker) and cabbage (*X. campestris* pv. *campestris* causing black rot)¹⁷ (Fig. 1b, and Supplementary Table 3). These observations indicate that this system might play important roles in supporting a successful infection and colonization, which led us to investigate in depth the molecular mechanisms governing XyG processing by *Xanthomonas* and its potential biological functions.

XyGUL endo-enzyme exploits arginine-carbohydrate interactions

The first enzymatic unit of the *Xanthomonas* XyGUL is encoded by XAC1770 (named here as *XacXeg74*) and belongs to the GH74 family, which is known to have high specificity for XyGs^{18,19}. According to kinetic characterization and cleavage pattern analysis, *XacXeg74* is an endo-dissociative enzyme generating a broad distribution of xyloglucan oligosaccharides (XyGOs), but preferentially Glc4- and Glc3-based products (Fig. 2a, and Supplementary Fig. 3c). These results indicate that *XacXeg74* would accept both X (α -D-Xylp-(1,6)- β -D-Glcp-(1-)) and G (-4)- β -D-Glcp-(1-) motifs at the -1 subsite, correlating with the presence of a glycine residue (G465) that confers such capacity to group I members of the GH74 family¹⁹ (Fig. 2b).

To get further insights into XyG recognition by the *Xanthomonas* GH74 enzyme, the crystal structure of *X. campestris* pv. *campestris* GH74 enzyme (*XccXeg74*, sharing 84% of sequence identity with *XacXeg74*) was determined in complex with the disaccharide XG spanning the subsites +1 and +2 (Supplementary Fig. 3a,b, and Supplementary Tables 4 and 5). Structural comparisons with the closest structurally characterized xyloglucanases from *Caldicellulosiruptor lactoaceticus* (*ClGH74A19*, PDB ID: 6P2M and SeqID 39.8%) and *Niastella koreensis* (*NkGH7419*, PDB ID: 6P2L and SeqID 34.2%) revealed remarkable differences in the molecular basis for substrate recognition. *ClGH74A* and *NkGH74* rely on several CH- π interactions for substrate anchoring, including at least 4 aromatic residues at the

subsites -3, +1, +3 and +5, and possibly two additional aromatic residues forming the subsites -4 (W126) and -5 (W599) in *ClGH74A* (Fig. 2b). However, except for the -3 (Y128) and the +1 (W396) aromatic-based subsites, the other aromatic residues are absent in *Xanthomonas* GH74 enzymes, in particular at the +3 and +5 subsites known to be critical for the endo-processive mode of action¹⁹ (Fig. 2b,c). The lack of aromatic residues in the subsites +3 and +5 of *XccXeg74* is, therefore, consistent with its endo-dissociative mode of action on XyG.

XccXeg74 contains three conserved motifs that could compensate for the lack of aromatic platforms in specific subsites: two arginine residues (R132 and R627) at the -4 subsite, a β -hairpin (residues A211-G226) inserted in the η 2- β 14 loop at the N-terminal lobe, and an arginine cluster at the positive subsite region (Fig. 2b-e, and Supplementary Fig. 4). The β -hairpin interacts concurrently with both galactosyl and xylosyl decorations at the -2 position and with the +1 xylosyl moiety via stacking contacts with H218 (Fig. 2d). In addition, the arginine residue (R510) stacks with the XyG backbone at the +2 position, and the other three nearby conserved arginine residues (R354, R356, and R513) are strategically located to establish polar and/or stacking interactions with the saccharide at the +3 subsite (Fig. 2e). This molecular strategy of carbohydrate recognition based on arginine residues observed in *Xanthomonas* GH74 enzymes is a distinguishing feature among GH families²⁰, which typically rely on aromatic CH- π interactions for carbohydrate binding.

The uptake of XyGOs is mediated by TonB-dependent transporters

After the extracellular cleavage of XyG backbone, the following reactions for the deacetylation and breakdown of the released oligosaccharides likely occur at the periplasm, either by the action of free or membrane-anchored enzymes, as indicated by signal peptide analysis and subcellular localization predictions (Supplementary Table 1). Supporting this hypothesis, the knockout of both TonB-dependent transporters (TBDTs) encoded by the XyGUL (XAC1768 and XAC1769) was highly detrimental to *X. citri* growth with XyGOs as carbon source, but not to the growth with a mixture of its monosaccharides, indicating that depolymerization of XyGOs occurs after passing the outer-membrane using specific TBDTs. Individual knockout of these transporters supports a major role for XAC1769 in XyGOs uptake. In the absence of XAC1768, XAC1769 was sufficient to maintain the bacterial growth akin to the wild-type strain using XyGOs as carbon source, but the opposite was not observed. The knockout of XAC1769 impaired the growth in the late log phase, indicating the importance of

this transporter as the XyGOs concentration decreases in the culture medium (Supplementary Fig. 5).

XyGOs deacetylation by *Xanthomonas* involves a distinctive acetyltransferase

Xanthomonas XyGUL harbors a putative esterase gene (XAC1771) initially annotated as sialate 9-O-acetyltransferase due to the low similarity with current known carbohydrate esterase families (Supplementary Table 2). The presence of an esterase is in accordance with the fact that fucosylated XyGs from dicotyledons can be acetylated at O6 position of the galactosyl moiety²¹. Acetylation modifies the physical and chemical properties of carbohydrates, limiting enzyme accessibility. Therefore, acetate removal is a key step toward efficient processing of this polysaccharide by downstream glycoside hydrolases.

The *Xanthomonas* XyGUL esterase, named here *XacXaeA* (xyloglucan acetyltransferase), showed high activity on *p*NP-acetate and did not accept moieties longer than acetyl as substrates (Supplementary Tables 6-8, and Supplementary Fig. 6h). It is specific for O-acetylation since it was not capable of cleaving N-acetylated carbohydrates (Supplementary Fig. 7). However, *XacXaeA* showed activity on a broad range of O-acetylated mono- and disaccharides and did not show a positional preference for acetylated oxygens (Supplementary Fig. 7). As expected, *XacXaeA* was active towards cell wall extracted xyloglucan oligosaccharides, deacetylating distinct types of structures such as XXLG/XLXG, XXFG and XLFG (Supplementary Figs. 8 and 9, and Supplementary Table 9).

To get insights into the modular structure and molecular determinants for O-acetyl transferase activity, the crystallographic structure of *XacXaeA* was solved using Zn²⁺ single-wavelength anomalous dispersion (SAD) (Supplementary Table 4). The catalytic domain displays the SGNH hydrolase fold (Fig. 3a-c) found in several lipases as well as esterases²² from families CE223, CE324, CE625, CE1226 and CE1727 (Fig. 3d). However, it is composed of two halves (residues 104-216 and 397-541) due to the insertion of a domain (residues 217-396, named X448 in the CAZy database) in the $\alpha 5$ - $\eta 3$ loop (Fig. 3a-c). Notably, both N- (residues 24-103) and C-terminal (residues 542-638) extensions exhibit an antiparallel seven-stranded β -sandwich fold that did not resemble any known domain at the sequence level. These two iso-structural β -sandwiches are intimately linked to the esterase core, forming a monolithic structure (Fig. 3b,c). Such structural architecture diverges from carbohydrate esterase (CE) families described in the CAZy database so far.

The active site encompasses the classical catalytic triad (Asp-His-Ser) (Fig. 3a, Supplementary Figs. 10a and 11), reminiscent from proteases^{28,29}. In addition, it conserves the electropositive oxyanion hole (Supplementary Fig. 10a-c), an ancestral and recurrent feature of enzymes from the GDSL and GDSL-like families of esterases and lipases^{30,31}. Interestingly, the catalytic triad is imprinted on a flat surface (Fig. 3c, and Supplementary Fig. 10b) that is uncommon in other known CE families described in the CAZy database²⁰ (Supplementary Fig. 10d-h). This observation agrees with a lack of selectivity to O-acetylated simple sugars since it does not seem to impose steric penalties to any C5 or C6 mono- and di-saccharides. Although both N- and C-terminal β -sandwich domains are remote to the catalytic center, it is proposed here that they might serve as an extended platform for XyGOs anchoring during acetate removal. Moreover, the internal X448 domain, which was not observed in the crystallographic structure, probably adds another component in the recognition mechanism of complex substrates by this new type of carbohydrate acetyltransferase. These functional and structural results allow the classification of *XacXaeA* as the founding member of the CE20 family.

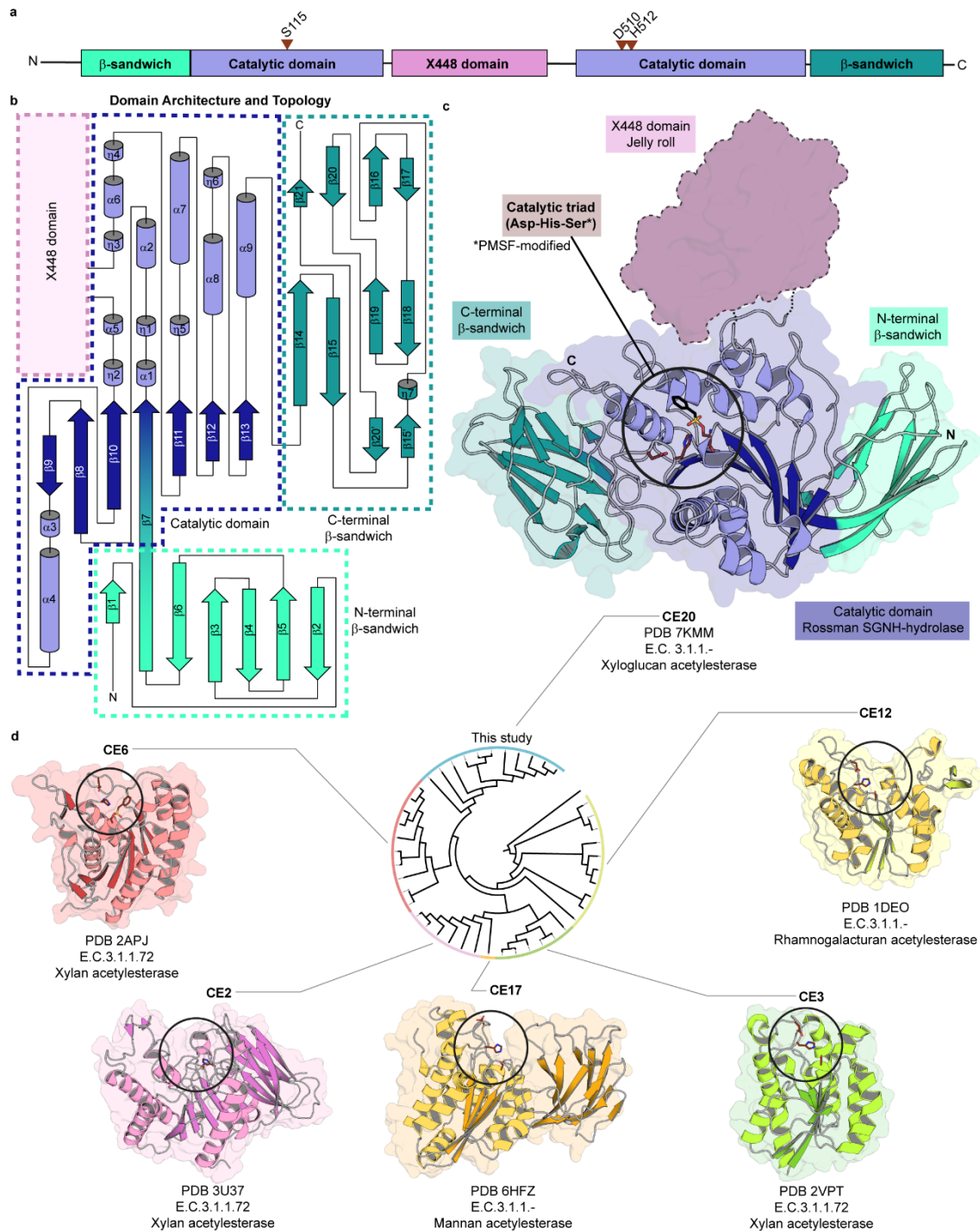


Fig. 3: A distinctive type of carbohydrate acetyltransferase. a, domain organization showing the position of catalytic residues (red triangles); b, structural topology and c, crystal structure color-coded according to panel a. d, Dendrogram of CE families based on phylogenetic analysis of the catalytic domain showing the structure of a representative member of each family. Circles indicate the active site. Catalytic residues (carbon atoms in brown) and PMSF (carbon atoms in black) are shown as sticks. The *Xanthomonas* acetyltransferase, discovered in this study, is the founding member of the CE20 family.

***Xanthomonas* enzymatic cascade for XyGOs breakdown**

The α -1,2-L-fucosidase activity (EC number 3.2.1.63) was biochemically observed in a GH95 member of the XyGUL (XAC1774), named here *XacAfc95*. Despite more than one activity has been reported for this family, *XacAfc95* appeared to be very specific to L-fucose, which is equivalent to 6-deoxy-L-galactose (Supplementary Tables 6-8, and Supplementary Fig. 12h). Its crystal structure (Supplementary Table 4) conserves the canonical domain architecture of the GH95 family that consists of an N-terminal β -supersandwich (residues 36-278), α -toroidal six-hairpin catalytic domain (residues 349-703), and a C-terminal β -sandwich (residues 704-790) (Fig 4a). The catalytic domain is connected to the supersandwich domain by a helix-rich linker (residues 279-348) and this multi-domain protomer forms dimers in solution (Supplementary Figs. 13j-l and 14, and Supplementary Table 10). Structural comparisons indicate that the general bases N397/N399 (carboxylate-activated) and the general acid D690 are conserved in relation to the only two structures available for this family, the α -L-galactosidase from *B. ovatus*³² (BACOVA_03438, PDB ID 4UFC, SeqID 42.88%) and the α -1,2-L-fucosidase from *Bifidobacterium bifidum*³³ (*BbAfcA*, PDB ID 2EAB, SeqID 30.36%) (Fig. 4b).

Previous comparisons between BACOVA_03438 (α -L-galactosidase) and *BbAfcA* (α -1,2-L-fucosidase) led to the suggestion that the only polymorphic position at the -1 subsite would confer specificity to L-galactose or L-fucose, although the authors also pointed that a conclusive inference of the functional relevance of this polymorphic residue is hindered by the very limited structural data available for this family so far³². Based on structural analyses, they proposed that the presence of a threonine at this position would allow a hydrogen bond with the L-galactose O6 atom, whereas a histidine would contribute to aliphatic interactions with the L-fucose C6 methyl group. However, *XacAfc95*, which shares nearly 40% sequence identity with characterized GH95 α -1,2-L-fucosidases involved in XyG depolymerization^{12,20}, contains a threonine at the referred position and showed high specificity to L-fucose, contraposing the initial role proposed for this residue as a determinant for L-galactose preference (Fig 4b, Supplementary Tables 6-8, and Supplementary Fig. 12h). In addition, the mutation T395H did not result in any change of substrate preference, supporting a less relevant role of this polymorphic position for specificity in the GH95 family (Supplementary Fig. 15a). Besides *XacAfc95*, another characterized GH95 α -1,2-L-fucosidase (*Blon_2335*)³⁴ conserves a threonine at this polymorphic position, corroborating this hypothesis (Supplementary Fig. 16).

These findings point to a more elaborate mechanism of substrate selectivity in the GH95 family that is not limited to direct interactions with the residues forming the -1 subsite.

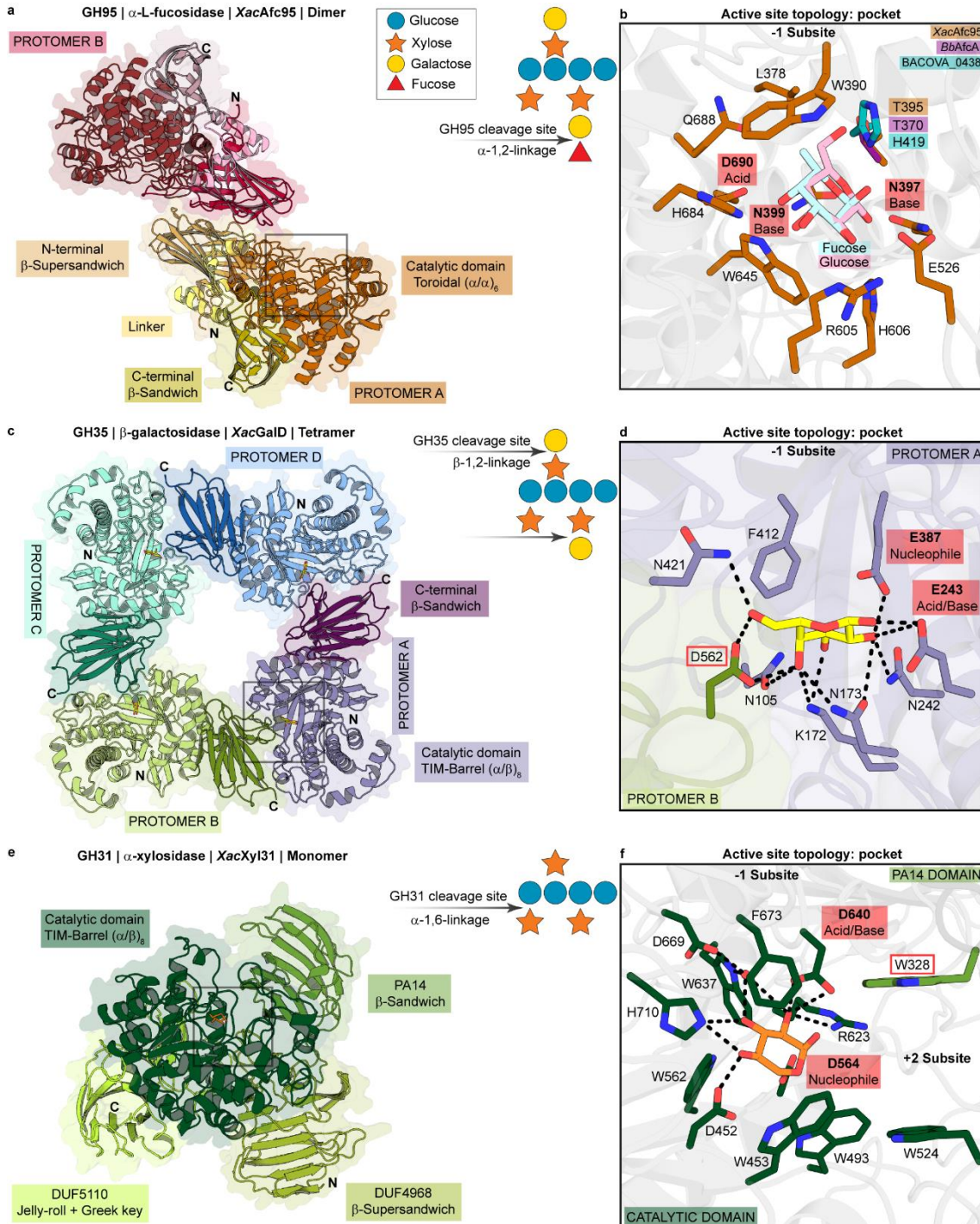


Fig. 4: Molecular basis for the enzymatic cleavage of XyGOs decorations in *Xanthomonas*. a, c, e, Crystallographic structures of the enzymes that release the substitutions of XyGOs (diagram where blue circle = glucose, orange star = xylose, yellow circle = galactose and red triangle = fucose), highlighting their domains in shades of a specific color defining each protomer or monomer as indicated in the figure.

b, d, f, Amplified view of the enzymes active sites indicated with boxes in panels a, c, e. Residues from the -1 subsite and ligands are shown as sticks (Glucose = pink carbon atoms, Fucose = blue carbon atoms, Galactose = yellow carbon atoms and Xylose=orange carbon atoms). The red outline indicates residues selected for mutational analysis. Catalytic residues are indicated (pink boxes).

The subsequent substitution to be cleaved after the α -1,2-L-fucoside removal is a β -1,2-galactosyl moiety. According to biochemical characterization assays, this step is performed by a β -galactosidase belonging to the GH35 family (XAC1772) with functional dependence to oligomerization. This enzyme is orthologous to GalD from *X. campestris* pv. *campestris*³⁵ and forms tetramers in solution (Supplementary Fig. 17a). Structural analysis revealed that the oligomerization interface involves the catalytic domain (TIM-barrel fold, residues 75-458) of one protomer and the accessory β -sandwich domain (residues 459-585) from the other subunit (Fig. 4c, and Supplementary Fig. 18). Since the functional relevance of oligomerization has not been investigated for this family so far, mutations were designed to address this question. The insertion of an arginine residue (S106R mutation) at the oligomeric interface resulted in stable monomers in solution (Supplementary Fig. 17b), which were devoid of catalytic activity (Supplementary Fig. 15b), supporting a critical role of tetramerization to the function of *Xanthomonas* GH35 members.

The crystal structure of *Xac*GalD with D-galactose (Supplementary Fig. 19a,b, and Supplementary Tables 4 and 5) revealed that residues in the oligomerization interface also interacts with the saccharide (Fig. 4d), demonstrating that the active-site pocket is completed by residues from the vicinal protomer. Alanine mutation of the residue D562 from the β -sandwich domain from the neighboring protomer, which interacts directly with the D-galactose in the active-site pocket (D562A), drastically reduces the enzyme substrate affinity (Supplementary Fig. 15b). Interface analysis of structurally characterized members of the GH35 family, including from archaea, eukaryota, and distinct bacterial phyla (Bacteroidetes, Proteobacteria and Firmicutes) points out that functionalization by tetramerization is a conserved feature across the phylum Proteobacteria (Supplementary Table 11, and Supplementary Fig. 20). Besides *Xanthomonas* GH35 enzymes, other representatives from this phylum in the CAZy database, such as CC0788 from *Caulobacter vibrioides* and *Cj*Bgl35A from *Cellvibrio japonicus*, also form stable tetramers in a similar fashion as *Xac*GalD, which contrasts to the quaternary structures observed in other phyla (Supplementary Table 11).

The next step in the cascade of XyGOs processing involves an α -xylosidase from the GH31 family, according to biochemical analysis of the enzyme encoded by XAC1773 (Supplementary Tables 6-8), named here *XacXyl31*. The crystal structure of this enzyme (Supplementary Tables 4 and 5) revealed a conserved active-site pocket compared to other bacterial α -xylosidases from this family, including the catalytic residues (D564 as nucleophile and D640 as acid/base) and the aromatic residue (W453). The latter introduces a steric barrier to C6 saccharides, favoring only the accommodation of C5 sugars such as xylose at the -1 subsite³⁶ (Fig. 4e, f, and Supplementary Fig. 19c, d).

Another conserved feature of GH31 α -xylosidases specific to XyGOs present in *XacXyl31* is the four-domain arrangement consisting of a central TIM-barrel catalytic domain (residues 392-757) and three all- β fold accessory domains (DUF4968, residues 38-219 and 372-391; DUF5110, residues 758-957; and PA14, residues 220-371)^{36,37} (Fig. 4e). These domains comprise a monolithic tertiary structure and do not establish oligomeric contacts in any of the GH31 members structurally characterized so far, including the enzyme reported here (Supplementary Fig. 13g-i, and Supplementary Table 10). Among the four accessory domains, the PA14 is the only participating in the active site interface (Fig. 4e, f)³⁷⁻³⁹. This domain introduces an aromatic platform (W328) at the +2 subsite that along with W524 from the TIM-barrel catalytic domain establishes stacking interactions with XyGOs backbone (Fig. 4f)^{37,40,41}. The residue W328 is highly conserved in GH31 α -xylosidases and the mutation W328A was detrimental to the catalytic activity (Supplementary Fig. 15c), supporting the importance of this region for α -xylosidase activity in the GH31 family.

As previously described in *C. japonicus*, GH31 α -xylosidases active on XyGOs may cleave specifically α (1 \rightarrow 6)-Xylp moieties appended to the non-reducing end of the backbone, requiring the cooperation of XyGOs-specific β -glucosidases to complete XyGOs depolymerization^{37,42}. *XacXyl31* also presented this specificity (Supplementary Fig. 21), but no β -glucosidase-encoding gene was found in the *Xanthomonas* XyGUL.

On the other hand, five potential β -glucosidases are present in the *X. citri* genome, all belonging to the polyspecific GH3 family. The heterologous expression and biochemical characterization of these enzymes revealed that two of them are β -xylosidases (*XacXyl3A* - XAC3076 and *XacXyl3B* - XAC4231) and the other three are β -glucosidases (*XacBgl3A* - XAC1448, *XacBgl3B* - XAC1793 and *XacBgl3C* - XAC3869) (Supplementary Tables 6 and 7).

These three β -glucosidases were expressed in the presence of XyGOs with higher expression levels of *XacBgl3C* and *XacBgl3B* compared to *XacBgl3A* (Supplementary Table 12) (see section below). In addition, all three enzymes were capable of releasing the non-reducing glucosyl moiety from the XyG-derived oligosaccharides GXXG and GXG, which are the products of *XacXyl31* using XXXG and XXG as substrates, respectively (Supplementary Figs. 21 and 22).

XacBgl3B displayed the highest activity on GXXG and GXG substrates and is predicted to be an outer membrane associated protein, probably exposed to the periplasm (Supplementary Fig. 22, and Supplementary Table 1). *XacBgl3A*, although predicted to be periplasmic, seems to be more specific to cleave β -1,3-glucooligosaccharides instead of XyGOs-derived β -1,4-glucooligosaccharides (Supplementary Figs. 23 and 24, and Supplementary Tables 1 and 8). *XacBgl3C* displayed a more generalist substrate profile and is predicted to be cytoplasmic, indicating that it might support the final steps of glucooligosaccharides cleavage coming from different sources (Supplementary Figs. 23 and 24, and Supplementary Tables 1 and 8).

Together, these findings point to *XacBgl3B* as being the major β -glucosidase to alternate with *XacXyl31* on the breakdown of XyGOs intermediates until reaching the final substrate β -1,4-glucobiose. At this last step, *XacBgl3B* seems also to play an important role, since it was the most efficient on β -1,4-glucobiose cleavage, compared to *XacBgl3A* and *XacBgl3C* (Supplementary Fig. 24, and Supplementary Table 8).

XyG depolymerization products activate the expression of virulence factors

The presence of XyGULs in both mesophyllic and vascular *Xanthomonas* pathogens indicates that XyG processing might play a role in bacterial colonization, pathogenicity and/or survival in host plants. To explore this hypothesis, the global gene expression profile of *X. citri* was assessed in the presence of XyGOs, which resulted in 276 differentially expressed genes among its 4281 protein coding sequences (Fig. 5a, and Supplementary Data 1).

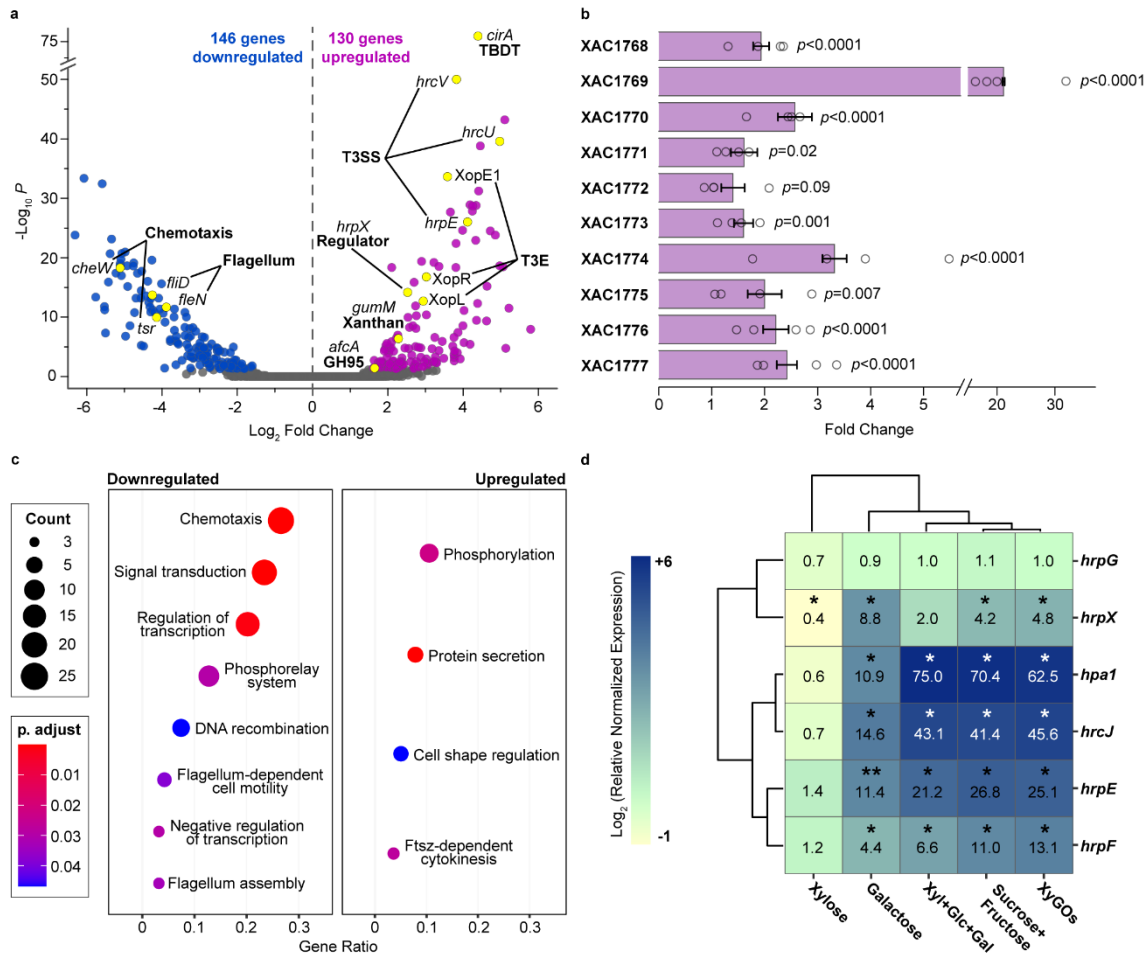


Fig. 5: Transcriptional responses to XyGOs in *X. citri*. a, Volcano plot of RNA-seq data highlighting differentially expressed genes (DEGs; blue circles = downregulated and purple circles = upregulated). In yellow are DEGs acting on processes important for virulence or XyG utilization. Genes were considered differentially expressed according to Wald test implemented in DESeq2. p-values were adjusted for multiple tests using the Benjamini-Hochberg (BH) method implemented in DESeq2. Thresholds: p-adjusted < 0.05 and |log₂ Fold Change| > 1. b, Transcription levels of XyGUL genes, downstream genes encoding xylose metabolism and MFS transporter in RNA-seq experiments. Data shown as mean ± SE from four biologically independent experiments (n=4) (open circles). Threshold: BH adjusted p-value < 0.05. c, Gene ontology (GO) enrichment analysis of DEGs. Circles size and color represent the counts and BH adjusted p-values, respectively, according to the legend at the left. Gene ratio is the number of DEGs related to a GO term divided by the total number of annotated DEGs. Categories were considered enriched based on hypergeometric test, implemented in the clusterProfiler 3.14.3 R/Bioconductor enrich function⁶⁰. d, Clustered heat map of RT-qPCR data of *hrp* genes color-coded according to the log₂ (-1 to +6) of the mean normalized expression relative to the control (numbers inside the boxes). Two-tailed t-test: *, p<0.001; **, p<0.0001; Xyl = xylose, Glc = glucose and Gal = galactose. See details in Supplementary Fig. 25.

As expected, the presence of XyGOs in the medium increased the expression of XyGUL genes (XAC1768-XAC1774), downstream genes related to xylose metabolism (XAC1775 and XAC1776) and an MFS transporter (XAC1777), supporting the relevance of this system for XyG processing in vivo (Fig. 5b). Moreover, these oligosaccharides up-regulated two GH43 encoding genes (XAC1275 and XAC4183) that are not in the vicinity of the XyGUL (Supplementary Data 1). The enzyme encoded by XAC4183, named here *XacAbf43A*, belongs to an underexplored subfamily, the GH43_9, with only one reported member characterized so far with a weak arabinofuranosidase activity⁴³. To evaluate this activity in the *Xanthomonas* member of this subfamily, *XacAbf43A* was produced, and biochemical assays confirmed its α -L-arabinofuranosidase activity (Supplementary Tables 6 and 7). The recombinant production of XAC1275 did not yield a soluble and stable protein; however, it displays 37% of sequence identity with the GH43_12 arabinofuranosidase from the *B. ovatus* XyGUL⁴⁰, pointing to a potential similar functional specificity. The fact that these GH43 genes are conserved across *Xanthomonas* bacteria, positively regulated by XyGOs and functionally related to orthologues present in *Bacteroides* XyGUL indicate that *Xanthomonas* could act on both types of XyGs, fucogalactoxyloglucans (present in plants from the Rosids clade⁴⁴) and arabinoxyloglucans (present in plants from the Asterids clade, in the Solanaceae and Oleaceae families⁴⁵).

The sensing of XyG depolymerization products by *Xanthomonas* also modulated other aspects of the bacterial metabolism (Fig. 5c). While repressing genes related to chemotaxis and flagellar motility, XyGOs stimulated processes that play essential roles in the early stages of plant infection⁴⁶, including xanthan gum biosynthesis and secretion of type III protein effectors into the host cells. Seven gum genes (XAC2574-XAC2580), the entire T3SS cluster (XAC0393-XAC0417), and 14 effector protein genes were up-regulated by XyGOs (Supplementary Data 1, and Supplementary Table 13), indicating an unprecedented role of XyG depolymerization products in bacterial virulence. As XyGs have a complex structure, RT-qPCR assays were performed to assess whether the activation of T3SS genes depends on the structure of XyGOs or its basic core constituents, i.e., glucose, galactose, and xylose. In these experiments, the mixed-sugar condition (glucose plus galactose and xylose) activated the expression of T3SS genes akin to the XyGOs condition, showing that the signaling effect of XyGOs relies on its monomers and not on its complex structure (Fig. 5d, and Supplementary Fig. 25).

To gather insight into the signaling pathway assessed by XyG depolymerization products, we also evaluated the expression of master regulators of T3SS expression, *hrpG* and

*hrpX*⁴⁷. Although *hrpG* was not activated in any of the tested conditions, the expression of *hrpX* increased in the presence of galactose, XyGOs and sugar mix, being maximal in the galactose condition (Fig. 5d, and Supplementary Fig. 25). This result shows that galactose is sufficient for the transcriptional activation of *hrpX*. However, for all T3SS genes analyzed, the transcription was further stimulated when galactose was supplemented with glucose and xylose, indicating that a second signal coming from other XyG components boosts the expression of T3SS genes (Fig. 5d, and Supplementary Fig. 25).

The combination of the monosaccharides mimicking the XyG breakdown products can also be achieved by the synchronized depolymerization of other plant polysaccharides and is as potent as the combination of sucrose and fructose in stimulating the T3SS expression in vitro (Supplementary Fig. 25), supporting that *Xanthomonas* relies on multiple and redundant sources of signals to trigger virulence and modulate host responses. Therefore, to confirm this redundancy hypothesis, we evaluated whether *X. citri* pv. *citri* would maintain its virulence even with the knockout of XyGUL genes (TBDTs and xyloglucanase) or an adjacent MFS transporter (Supplementary Fig. 26). As expected, the wild-type phenotype was preserved in these mutants, supporting that the activation of T3SS by XyG depolymerization products is probably compensated via functional redundancy or via alternative pathways for virulence activation, especially in the case of Δ XAC1768-69 deletion. The in vitro growth of this mutant in minimal medium containing XyGOs was severely reduced, indicating that the lack of XAC1768-69 genes impairs the uptake of these oligosaccharides (Supplementary Fig. 5a).

The deletion of the inner-membrane MFS transporter gene (XAC1777) did not affect the bacterial growth using either XyGOs or its basic components as primary carbon source, supporting that at least one of the other 40 MFS transporters encoded by *X. citri* genome compensate its absence (Supplementary Fig. 5, and Supplementary Table 14). Notably, the strain lacking the GH74 xyloglucanase (Δ XAC1770) displayed a XyG depolymerization halo similar to the wild-type strain, suggesting that another endo- β -1,4-glucanase would be functionally redundant to the XyGUL GH74 enzyme (Supplementary Fig. 27). A search for endo- β -1,4-glucanases in *X. citri* genome resulted in one GH9 member (XAC2522, *XacEgl9*), one GH8 member (XAC3516, *XacCel848*), and five putative GH5 glucanases (XAC0612 – subfamily GH5_1, *XacEngXCA*; XAC0028 – subfamily GH5_5, *XacEgl5A*; XAC0029 – subfamily GH5_5, *XacEgl5B*; XAC0030 – subfamily GH5_5, *XacEgl5C*; and XAC0346 – not yet assigned to a subfamily)⁴⁹. The recombinant production and activity assays of these enzymes revealed that only *XacEgl9*, *XacEgl5B*, and *XacEngXCA* are able to cleave XyG, with

XacEgl9 showing the highest specific activity on this polysaccharide (Supplementary Table 15). This result indicates that these glucanases might compensate the absence of the GH74 xyloglucanase in the Δ XAC1770 mutant. Characterization of *XacEgl9* revealed kinetics parameters on XyG akin to those found for *XacXeg74* (Supplementary Table 8, and Supplementary Figs. 6g and 12i), supporting the role of other endo- β -1,4-glucanases in the XyG cleavage in strains lacking the GH74 enzyme.

Discussion

Here, we show that most *Xanthomonas* species, a highly diverse bacterial genus that infects hundreds of plants, have an intricate enzymatic toolbox to break down XyGs. The *Xanthomonas* XyGUL encodes oligomeric and multi-modular glycoside hydrolases (GH74 xyloglucanase, GH31 α -xylosidase, GH35 β -galactosidase and GH95 α -L-fucosidase) and a distinguishing carbohydrate acetylsterase with no significant similarity with known CAZy families and not present in any similar XyGUL characterized so far^{5,6,12} (Figs. 1 and 6). This novel acetylsterase, *XacXaeA*, is the founding member of the CE20 family.

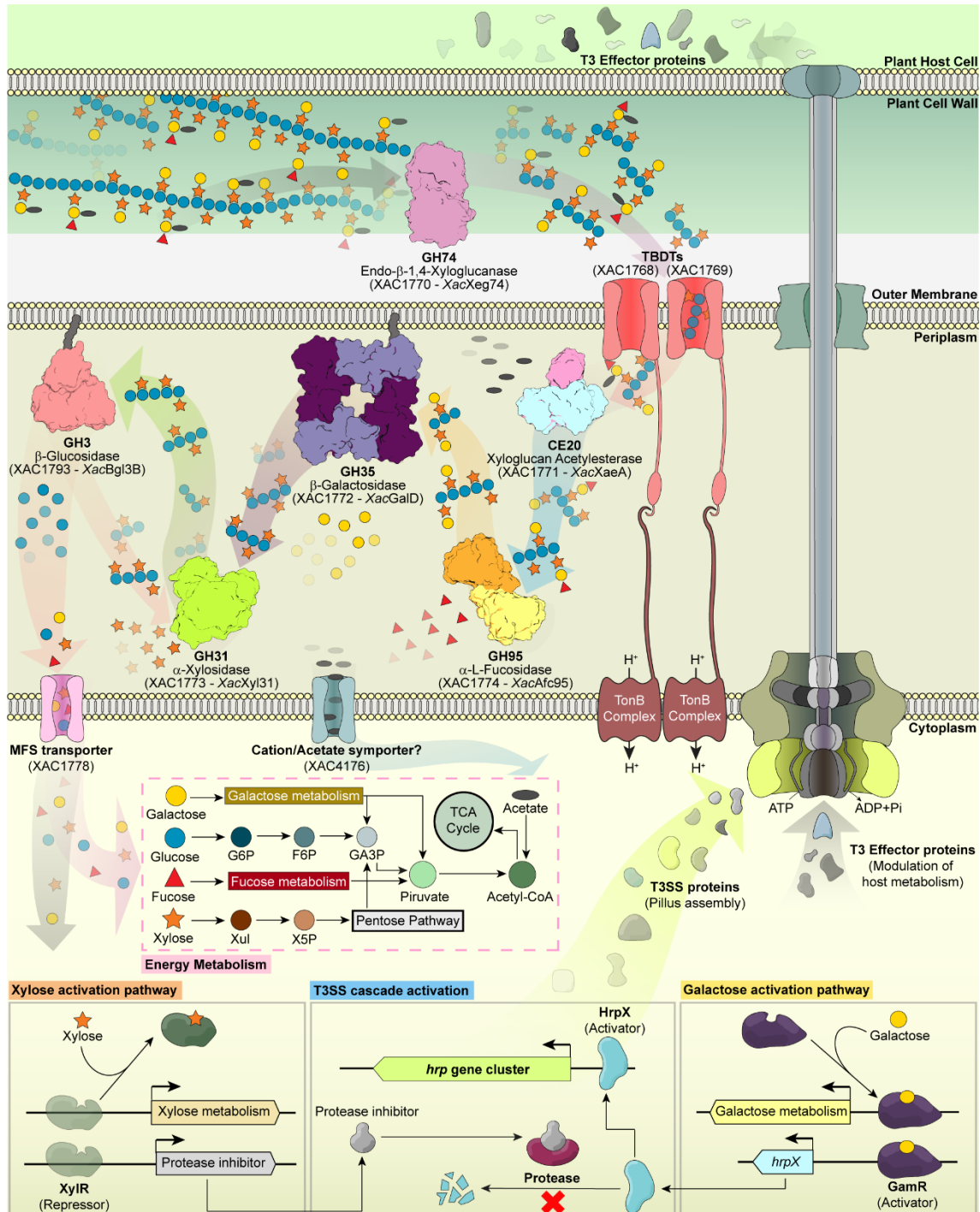


Fig. 6: XyG enzymatic breakdown and the role of released products in the metabolism and virulence of *X. citri*. Representation of the pathogen-host cell interface highlighting how XyG is progressively depolymerized by *Xanthomonas* enzymes and internalized by transmembrane transporters. In this bacterium, monosaccharides released from XyG induce the expression of multiple genes, including the T3SS machinery via mechanisms suggested based on transcriptional data and literature⁵⁴⁻⁵⁶ (details in Supplementary Fig. 28, and Supplementary Table 1). Abbreviations: T3, Type III; T3SS, Type III secretion system; TBBDTs, TonB-dependent transporters; MFS, major facilitator superfamily transporter;

G6P, D-glucose-6-phosphate; F6P, D-fructose-6-phosphate, GA3P, D-glyceraldehyde 3-phosphate; Xul, D-xylulose; X5P, D-xylulose 5-phosphate; and TCA, Tricarboxylic Acid cycle.

Transcriptomic, biochemical and gene deletion analyses revealed that the *Xanthomonas* XyGUL is complemented by other conserved CAZyme genes encoding GH9 and GH5 enzymes with xyloglucanase activity, GH3 β -glucosidases and GH43 arabinofuranosidases. The GH9 (*XacEgl9*) and GH5 (*XacEgl5B*) enzymes serve as redundancy components to the pivotal xyloglucanase activity that initiate XyG breakdown and the GH43 enzymes, such as *XacAbf43A*, confer the ability to these bacteria to cleave both fucogalactoxyloglucans⁴⁴ and arabinoxyloglucans⁴⁵. In the final steps of XyG processing, the GH31 α -xylosidase (*XacXyl31*) acts coordinately with a XyGOs-active β (1 \rightarrow 4)-glucosidase (*XacBgl3B*) at the non-reducing end of XyGOs, cycling between xylosyl side chain cleavage and glucosyl main chain removal, as previously observed in the saprophyte *C. japonicus*^{5,42}. A difference is that *C. japonicus* displays a β -glucosidase with the highest efficiency toward the intermediate GXXG and another one with the highest efficiency on β -1,4-glucobiose, whereas in *X. citri* the enzyme *XacBgl3B* seems to play both roles. These observations highlight that microbial systems, according to their ecological niches, have evolved singular and equally complex molecular strategies to cope with the structural and chemical diversity of XyGs.

From a mechanistic point of view, the enzymes encoded by *Xanthomonas* XyGUL also show distinguishing properties compared to other characterized homologues. The GH74 enzyme harbors an unusual arginine-based mechanism of substrate recognition, contrasting to the canonical strategy in the GH families based on aromatic CH- π interactions⁵⁰ (Fig. 2). The xyloglucan acetylerase features an unprecedented molecular architecture with two iso- β domains at both N- and C-termini and the insertion of an uncharacterized domain in the middle of the catalytic core (Fig. 3). The GH95 member demonstrates that substrate specificity in this family does not involve polymorphic positions in the -1 subsite, pointing to the relevance of indirect interactions in determining selectivity. In addition, the role of ancillary domains and oligomerization is prominent in the function of *Xanthomonas* XyGUL enzymes (Fig. 4).

Notably, the XyG depolymerization products play a role in the activation of multiple genes related to virulence in *Xanthomonas* spp., including those encoding effector proteins and the T3SS that inject these virulence factors into the host cells (Fig. 5). In the genus *Xanthomonas*, the activation of T3SS expression by carbohydrates such as sucrose and fructose in *X. campestris* pv. *campestris* and *X. campestris* pv. *vesicatoria*^{51,52}, and xylose in *X. oryzae*

pv. oryzae^{53–55}, has been previously demonstrated. However, the role of galactose in this process and XyG depolymerization as a source of T3SS inducers, are novel components in the complex regulatory mechanisms of virulence in these pathogens.

Our results demonstrate that galactose from XyG depolymerization activates the transcription of *hrpX*, a master regulator of T3SS expression (Fig. 6). The role of galactose as an inducer of *hrpX* gene correlates with a previous study in *X. oryzae* showing that activation of *hrpX* expression is mediated by a regulator of galactose metabolism, termed GamR⁵⁶ (XAC1767). In addition to galactose activation, a second signal from xylose, acting after the *hrpX* expression, likely contributes to the expression of T3SS genes. Similarly to that proposed for *X. oryzae*, the induction of T3SS by xylose might be associated with a post-transcriptional mechanism that suppresses HrpX proteolysis⁵⁴. Based on these observations, we propose that during XyG depolymerization, the released galactose activates the transcription of *hrpX*, likely by modulating the GamR activity, whereas the released xylose suppresses HrpX degradation, thus promoting the activation of T3SS genes and other HrpX-mediated processes (Fig. 6). It is noteworthy that most *Xanthomonas* bacteria are equipped with other CAZymes and polysaccharide utilization loci specialized in depolymerizing other hemicelluloses^{57,58} and pectins⁵⁹, which can also generate xylose and galactose, adding more layers of complexity in the modulation of virulence and pathogenesis mediated by host carbohydrate processing in these phytopathogens.

In conclusion, this work provides an in-depth understanding of the multi-enzymatic system employed by plant-associated bacteria for XyG depolymerization, and also establishes a novel component in the regulatory mechanisms of virulence and pathogenesis in *Xanthomonas*.

Methods

Phylogenetic analysis

Phylogenetic analysis of *Xanthomonas* species was performed based on a set of 92 single-copy core genes according to the UBCG pipeline 3.0⁶¹. Individual proteins were aligned separately using MAFFT 7.299b and then concatenated. Phylogenetic analysis was inferred using RAxML 8.2.0 and the PROTGAMMAWAG model with 1,000 bootstrap replicates. Applying a similar protocol, phylogenetic analysis of carbohydrate acetylsterases was

performed with sequences of the catalytic domain of characterized enzymes available at the CAZy database (<http://www.cazy.org/>) including homologous sequences of *XacXaeA*.

Molecular cloning and site-directed mutagenesis

The nucleotide sequence encoding the XyGUL and accessory enzymes were amplified from the genomic DNA of *X. citri* pv. *citri* 306 strain or *X. campestris* pv. *campestris* ATCC 33913 using standard methods (Supplementary Table 16) and confirmed by Sanger sequencing. PCR-amplified gene fragments were cloned into the pET28a and pETM11 vectors. The mutants were prepared using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) (Supplementary Table 17).

Protein expression and purification

The proteins were expressed in *Escherichia coli* strains as described in the Supplementary Table 18. Proteins were purified by two chromatographic steps as detailed in the Supplementary Table 19. Purified samples were analyzed by SDS-PAGE and dynamic light scattering (DLS) in a Malvern ZetaSizer Nano series Nano-ZS (model ZEN3600) instrument (Malvern Zetasizer). DLS data were collected and analyzed with Zetasizer (7.12) software to evaluate sample homogeneity.

Analytical hydrodynamic analysis

Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) experiments were performed at 25 °C using a triple-angle static light scattering detector miniDAWN™ TREOS and Optilab® T-rEX refractive index monitor (Wyatt Technology) coupled to an ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare) with a Superdex 200 HR 10/300 GL analytical size-exclusion column (GE Healthcare)⁶². 250 µL from purified wild-type *XacGalD* and mutant S106R at 50 µmol. L⁻¹ were injected separately into the column and eluted in 20 mmol.L⁻¹ Hepes pH 7.5, 150 mmol.L⁻¹ NaCl. Data were processed using the ASTRA V software 6.0 (Wyatt Technology).

Small-angle X-ray scattering (SAXS) data collection was performed with protein samples at different concentrations (1, 3, 5, and 10 mg. mL⁻¹) at the D01A-SAXS2 beamline at the Brazilian Synchrotron Light Laboratory (LNLS-CNPEM, Campinas, Brazil), using a CCD-Mar165 detector and fit2D software v. 18. Data were processed and analyzed with the ATSAS package 4.8.6⁶³ using the programs GNOM v. 5.0, DAMMIN v. 5.3, DAMAVER v. 5.0, CRY SOL v. 2.8.3, and SUPCOMB v. 2.3. SAXSMoW server⁶⁴ was used for protein molecular

weight determination, and oligomeric interface interaction energy was calculated using the PDBePISA server⁶⁵.

Protein crystallization, X-ray data collection and structure determination

Proteins were crystallized by the vapor diffusion method (Supplementary Table 20). Diffraction data were acquired under cryogenic conditions at the MX2 beamline from the Brazilian Synchrotron Light Laboratory (LNLS-CNPEM, Campinas, Brazil) using a Pilatus 2M detector (Dectris) and MXCuBE 2 (Qt4) software or at the BL9-2 beamline from the Stanford Synchrotron Radiation Lightsource (SSRL-SLAC, Menlo Park, USA) using a Pilatus 6M detector (Dectris) and BluIce 4.0 software. Data were indexed, integrated and scaled using the XDS package v. Jan 31st 2020 Built 20200417⁶⁶. *XccXeg74*, *XacGalD*, *XacXyl31* and *XacAfc95* structures were solved by molecular replacement method using the PHASER software from PHENIX package dev-3139⁶⁷ and the atomic coordinates from homologous proteins 2CN2⁶⁸, 4D1I⁵, 2XVG³⁷ and 4UFC³², respectively. Structure of *XacXaeA* was solved by zinc single-wavelength anomalous diffraction (SAD) using the programs SHELXC/D/E from CCP4i package 7.0.023⁶⁹. The initial model of *XacXaeA* was obtained with AutoBuild Wizard from the PHENIX package dev-3139⁷⁰ and further refined iteratively with COOT 0.8.9⁷¹ and PHENIX_refine dev-3139 programs. Structure validation was carried out with the Molprobtity server⁷². Metal-binding sites validation was performed with the CheckMyMetal server^{73–75}. Carbohydrate complexes structures were evaluated using Privateer software from CCP4i2 package 1.0.2 revision 5710⁷⁶ and figures were generated using Pymol v. 2.3 or 1.3. Data collection, processing and analyses are summarized in Supplementary Tables 4 and 5.

Glycoside hydrolase assays

XyGUL and accessory GHs activities were evaluated against several synthetic and natural substrates as described in Supplementary Table 21. The enzyme amount and reaction time for enzyme assays were determined based on linearity tests previously performed. Spectrophotometric data were collected in an Infinite 200 PRO microplate reader (Tecan) using the i-Control 1.10.4.0 software (Tecan). Kinetic parameters were determined from substrate saturation curves using the OriginPro 8.1 software. All enzyme assays consist of at least three independent experiments.

C. langsdorffii XyG was extracted from the cotyledons powder by washing three times with 80% ethanol solution at 80 °C for 10 min to remove low molecular weight carbohydrates. Polysaccharides were extracted by resuspending in water the ethanol-insoluble portion of the

material and incubating for 8 h at 80 °C under constant agitation. The mixture was filtrated, added 3 volumes of ethanol and centrifuged (12,000 g for 15 minutes). The supernatant was discarded and the pellet dried at 80 °C⁷⁷. The XyGOs released from *C. langsdorffii* XyG by *XccXeg74* enzyme were analyzed by the HPAEC-PAD system (Dionex) equipped with the CarboPac PA100 analytical column (Dionex). Identification of XyGOs was performed by MALDI-TOF on a Bruker Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics) in reflectron positive mode with a 19 kV voltage and covered mass within the m/z values of 700 to 3,500.

The activity of the enzymes *XacBglA*, *XacBglB*, and *XacBglC* on XyGOs was investigated after the removal of the xylosyl moiety at the non-reducing end glucosyl residue in XyGOs by the β -xylosidase *XacXyl31*. Reactions with *XacXyl31* (0.1 mg. mL⁻¹) were performed using 10 mg. mL⁻¹ XyGOs (Megazyme O-X3G4) incubated at 45 °C and pH 6.5. After 60 min, the reactions were stopped by boiling for 5 min and the final products (GXXG and GXG) were used for activity assays with the three β -glucosidases. *XacBglA*, *XacBglB*, and *XacBglC* reactions consisted of adding 0.1 mg. mL⁻¹ of each enzyme to the solution of GXXG and GXG. The temperatures chosen in the reactions were 45 °C (*XacBglA*), 25 °C (*XacBglB*) or 35 °C (*XacBglC*). Samples were collected at 0, 30 min, 1, 2, 3, and 24 h and reactions were stopped by the addition of methanol. A total of 5 μ L of the quenched reactions were added to 95 μ L of 5 μ mol. L⁻¹ xylohexaose (used as the internal standard) in water and injected into an LTQ XL TM linear ion trap mass spectrometer (Thermo Fisher Scientific) in scan mode (m/z 300-1,300). Samples were directly infused at a rate of 10 μ L.min⁻¹ into an ESI (+) source with a spray voltage maintained at 4.0 kV and heated to 250 °C in the source.

The xyloglucanase activity of wild-type and mutant *Xanthomonas* strains were monitored in plate assays. Cultures were grown overnight in LBON medium (1% m/v bacto peptone and 0.5% m/v yeast extract) at 30° C and 200 rpm, diluted to OD600 nm 0.4, plated (0.5 μ L) on solid NYG medium (5 g. L⁻¹ peptone, 3 g. L⁻¹ yeast extract, 20 g. L⁻¹ glycerol, 15 g. L⁻¹ agar) supplemented with 0.5% of tamarind xyloglucan (Megazyme), and grown for 40 h at 30 °C. Activity halos were revealed with 5 mg. mL⁻¹ Congo red and successive washes with 1 mol. L⁻¹ NaCl.

***Arabidopsis thaliana* cultivation and XyGOs enzymatic extraction**

Arabidopsis thaliana Col-0 seeds were sterilized in 70% ethanol solution for 2 min in a 2 mL tube. The supernatant was discarded and followed by addition of 10% (v/v) sodium

hypochlorite, 10 μL Tween 20 (Sigma-Aldrich) solution for 5 min under agitation. Then, the material was washed five times with sterile water and stratificated in water for 48 h at 4 °C protected from light. Seeds were plated over Murashige and Skoog⁷⁸ sterile media (Sigma-Aldrich) at half strength with 10% agar. Plates were exposed to a photosynthetic photon flux density of 200 $\mu\text{mol. m}^{-2} \text{ s}^{-1}$ for 2 h and then grew in a Phytotron chamber (Fitotron HGC Weiss Technik) for 6 days in the dark at 21 °C, 70% humidity. Seedlings (around 1 cm long in height) were harvested, weighted, frozen in liquid nitrogen and stored at -80 °C⁷⁹.

Frozen seedlings were homogenized in a Retschmill (model MM200, Retsch) at 25 Hz for 1 min. The grounded plants were washed three times by resuspending it in 1 mL methanol, vortexing, centrifuging at 10,000 g for 10 min and discarding the supernatant. The material was dried for 5 min under vacuum and washed twice with 500 μL of water, discarding the supernatant⁷⁹. The resulting residue was used for acetylated XyGOs extraction. A proportion of 1 μg of purified *XacXeg74* enzyme for every 50 mg of starting seedlings was utilized for digesting the acetylated XyG in 200 μL of 50 mmol. L⁻¹ ammonium formate pH 5.0 solution for 16 h at 30 °C, 450 rpm. Reaction was stopped by heating the mixture for 2 min at 95 °C. The resulting solution was centrifuged twice at 1,000 g for 5 min, the pellet discarded and the supernatant finally stored at -20 °C upon utilization.

Xyloglucan acetylsterase assays

Acetylated mono- and disaccharides were chemically prepared with excess of acetic anhydride in the presence of the catalyst pyridine⁸⁰ and validated by ¹H nuclear magnetic resonance. Spectra analysis showed that the acetylation on both monosaccharides (mannose pentaacetate, galactose pentaacetate, fucose tetraacetate, arabinose tetraacetate and xylose tetraacetate) and disaccharides (β -1,4-glucobiose octaacetate and saccharose octaacetate) was not selective. The ¹H nuclear magnetic resonance spectra were recorded on DD2 spectrometer (Agilent) from Brazilian Biosciences National Laboratory (LNBio-CNPq, Campinas, Brazil), operating in Larmor frequency of 499.726 MHz equipped with triple resonance probe. NMR data processing was performed using VnmrJ software (4.2 Revision A). Other used acetylated sugars were purchased including N-acetylglucosamine, N-acetylneuraminic acid (Sigma-Aldrich), α -glucose pentaacetate (Santa Cruz Biotechnology), β -galactose pentaacetate and β -glucose pentaacetate (Combi-Blocks). Reactions consisted of 0.01 mg. mL⁻¹ of the enzyme *XacXaeA* and 5 mmol. L⁻¹ of each acetylated sugars incubated during 15 min at 20 °C and 600 rpm in 50 mmol. L⁻¹ HEPES buffer pH 7.5. Reactions were stopped by adding 40 μL of methanol. Final products and residual substrates were monitored on a Waters Synapt HDMS

system at V mode, and ESI (+) with a spray voltage maintained at 3.0 kV and heated to 130 °C in the source using MassLynx 4.1 software. A volume of 15 μL of the quenched reactions and 2 μL of 1 mmol. L^{-1} xylotetraose (used as the internal standard) were added to 183 μL of water and injected into the mass spectrometer in scan mode (m/z 150–900) with direct infusion at a flow rate of 50 $\mu\text{L} \cdot \text{min}^{-1}$ ⁸¹.

Esterase reactions on acetylated XyGOs extracted from *A. thaliana* (see section above) consisted of incubating 0.02 mg. mL^{-1} of XacXaeA with the substrate (estimated concentration 5 to 10 $\mu\text{g} \cdot \text{mL}^{-1}$) for 2 and 24 h at 20 °C and 600 rpm on 50 mmol. L^{-1} HEPES buffer pH 7.5. 100 μL of quenched reactions were desalted using Oasis HLB cartridges (Waters). HLB cartridges were first activated with methanol and equilibrated with water according to the manufacturer's protocol. The samples were applied and then washed 7 times with 1 mL water. XyGOs were eluted with 150 μL 25% (v/v) methanol in water. Controls, final products and residual substrates were analyzed on a LTQ XL TM linear ion trap mass spectrometer (Thermo Fisher Scientific). The samples were directly infused at a rate of 10 $\mu\text{L} \cdot \text{min}^{-1}$ into the ESI (+) source in scan mode (m/z 150–2,000). The spray voltage maintained at 4.2 kV and heated to 280 °C in the source. CID-MS/MS fragmentation analysis of XyGOs were performed using different collision energies (15–35) and the isolation window was set to 1 Th. Estimated mass/charge ratio of acetylated and non-acetylated oligosaccharides were compatible to the literature^{44,79} and confirmed by MS/MS fragmentation fingerprint.

***Xanthomonas* cultivations**

For growth curve analysis, *X. citri* strains was cultured in LBON medium (1% m/v bacto peptone and 0.5% m/v yeast extract) containing 100 $\mu\text{g} \cdot \text{mL}^{-1}$ ampicillin at 30 °C and 200 rpm until mid-exponential phase. Then, the harvested cells were washed once and transferred to the modified minimal medium XVM2⁵² (XVM2m, without sucrose and fructose, containing different sugar sources at a final concentration of 5 mg. mL^{-1}), for an initial OD600 nm = 0.01. Growth was monitored for 30 h, at 30 °C, in a SpectraMax M3 Multi-Mode Microplate Reader (Molecular Devices). Four biological replicates were used for each condition. XyGOs used in *Xanthomonas* growth assays were prepared by incubating 5 mg. mL^{-1} tamarind xyloglucan (Megazyme) with XacXeg74 (4 $\mu\text{g} \cdot \text{mL}^{-1}$) at 30 °C for 14 h. The reaction was stopped by heating at 80 °C for 15 min.

RNA sequencing and analysis

Total RNA was extracted from 15 mL of *X. citri* cultures grown on XVM2m + XyGOs or XVM2m + glucose medium (see the section above) at the mid-exponential phase using the TRIzol/chloroform protocol⁸². Samples were further treated with RNase-free DNaseI (Invitrogen) and RNaseOUT (Invitrogen) and purified with the RNeasy Mini Kit (Qiagen), according to the manufacturer's recommendations. In addition, RNA integrity was evaluated in an Agilent 2100 Bioanalyzer (Agilent Technologies) and samples were quantified in a Qubit® 2.0 Fluorometer using the RNA BR assay kit (Life Technologies). Libraries were prepared according to the manufacturer's protocol of the TruSeq Stranded Total RNA kit (Illumina Inc.). Sequencing was performed on the Illumina HiSeq 2500 platform (LNBR-CNPEN, Campinas, Brazil). RNA-seq data were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE159288.

RNA-seq raw reads were filtered to remove low-quality reads and adapters sequences using Trimomatic v. 0.38⁸³ and rRNA reads were removed using SortMeRNA 2.1⁸⁴ (Supplementary Table 22). High-quality reads were mapped to the *Xanthomonas citri* pv. *citri* 306 genome⁸⁵ using Bowtie2 v.2.2.5 algorithm⁸⁶ and reproducibility among the biological replicates was assessed by the PCA (Principal Component Analysis) and Pearson correlation methods. Differential expression analysis was carried out by pairwise comparison between *X. citri* grown in XVM2m containing XyGOs and XVM2m glucose medium using $|\log_2 \text{Fold Change}| \geq 1$ and a p -adjusted ≤ 0.05 as thresholds using the Bioconductor DESeq2 v.1.18.1⁸⁷ package in the R v.3.4.1 platform⁸⁸.

RT-qPCR analysis

RNA-seq data were analyzed for the identification of potential reference genes. The arithmetic mean of the TPM values (transcription per million reads) of each gene was calculated in all conditions, followed by the determination of the variation coefficient and MFC (ratio between the maximum and minimum TPM value of each gene)^{89,90}. From the 20 potentially most stable genes, the targets XAC2293, XAC2177, XAC4047, and XAC4218 were selected for RT-qPCR experiments based on their mean TPM and p-values (Supplementary Table 23). The expression stability was evaluated for the potential reference genes and the Cq values (quantification cycles) were analyzed using three different statistical tools: BestKeeper⁹¹, NormFinder⁹² and RefFinder⁹³ (Supplementary Table 24, and Supplementary Fig. 29). RT-qPCR assays were performed in an Applied Biosystems ViiA™ 7 Real-Time equipment (Life

Technologies) using the Power SYBR® Green RNA-to-CT™ 1-Step Kit (Life Technologies) as detailed in Supplementary Table 25. The relative normalized expression (RNE) values for each gene were calculated according to the $2^{-\Delta\Delta C_t}$ method⁹⁴. Data were log-transformed and statistically compared by ANOVA and unpaired 2-tailed t-test using Prism 8.4.1 software (GraphPad). The correlation between gene expression data obtained in RNA-seq and RT-qPCR assays can be assessed in Supplementary Fig. 30.

Gene knockout in *X. citri*

Single (Δ XAC1768, Δ XAC1769, Δ XAC1770, Δ XAC1777) and double (Δ XAC1768-XAC1769) gene knockout mutants were obtained by a two-step allelic exchange procedure. DNA fragments (~1.2 kb) corresponding to regions upstream and downstream to the target genes were amplified by PCR from the *X. citri* genome (Supplementary Table 26). Each corresponding pair of fragments was ligated and then cloned into the pNPTS138 suicide vector⁹⁵ in the corresponding restriction sites. The plasmids were introduced into *X. citri* by electroporation (~2.3 kV, ~5 ms), and sucrose-sensitive and kanamycin-resistant colonies were selected (LBON-agar, 100 μ g. mL⁻¹ ampicillin, 100 μ g. mL⁻¹ kanamycin with and without 5% sucrose, respectively). This step selected colonies that suffered the first homologous recombination event, when the plasmid is inserted into the bacterial genome. These colonies were grown in LBON, 100 μ g. mL⁻¹ ampicillin without selection to allow the occurrence of the second recombinant event, when the plasmid is excised from the genome. The cultures were plated, and individual colonies were selected for simultaneous sucrose resistance and kanamycin sensitivity. Deletions were confirmed by PCR and DNA-sequencing (Supplementary Table 27).

Virulence assays

Plants of sweet orange (*Citrus sinensis* ‘Natal’) were infiltrated by the pinprick method⁹⁶ with water suspensions of *X. citri* at OD600 of 0.1 previously grown in LBON agar plates, supplemented with ampicillin (100 μ g. mL⁻¹), for 48 h at 28 °C. The assays were performed with three independent biological samples, each composed of 16 technical replicates. Plants were maintained under greenhouse conditions and monitored daily for the appearance of canker symptoms. Quantitative analyses of canker lesions were performed using ImageJ v. 1.53b.

Abbreviations

XyG, xyloglucan; XyGUL, xyloglucan utilization loci; CAZymes, carbohydrate-active enzymes; *Xac*, *Xanthomonas citri* pv. *citri*; T3SS, type III secretion system; T3E, type III effectors proteins; TBDTs, TonB-dependent transporters; GamR, galactose metabolism regulator; XyGOs, xyloglucan oligosaccharides; GH, glycoside hydrolase; Glc, glucose; *Xcc*, *Xanthomonas campestris* pv. *campestris*; SAXS, small angle X-ray scattering; CE, carbohydrate esterase; pNP, *para*-nitrophenyl; SAD, single-wavelength anomalous dispersion; PMSF, phenylmethylsulfonyl fluoride; SGNH, Serine-Glycine-Asparagine-Histidine; GDSL, Glycine-Aspartate-Serine-Leucine; DUF, domain of unknown function; MFS, major facilitator superfamily; RT-qPCR, real-time quantitative reverse transcription polymerase chain reaction; *hrp*, hypersensitive reaction and pathogenicity; GH74, glycoside hydrolase family 74; GH9, glycoside hydrolase family 9; GH3, glycoside hydrolase family 3; GH5, glycoside hydrolase family 5; GH35, glycoside hydrolase family 35; GH31, glycoside hydrolase family 31; GH95, glycoside hydrolase family 95; GH2, glycoside hydrolase family 2; GH43, glycoside hydrolase family 43; CE20, carbohydrate esterase family 20.

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes 7KN8 (*XccXeg74* complexed with XG oligosaccharide) [<https://doi.org/10.2210/pdb7KN8/pdb>], 7KMM (native *XacXaeA*) [<https://doi.org/10.2210/pdb7KMM/pdb>], 7KMN (native *XacGalD*) [<https://doi.org/10.2210/pdb7KMN/pdb>] , 7KMO (*XacGalD* complexed with galactose) [<https://doi.org/10.2210/pdb7KMO/pdb>], 7KMP (native *XacXyl31*) [<https://doi.org/10.2210/pdb7KMP/pdb>], 7KNC (*XacXyl31* complexed with xylose) [<https://doi.org/10.2210/pdb7KNC/pdb>] and 7KMQ (native *XacAfc95*) [<https://doi.org/10.2210/pdb7KMQ/pdb>]. RNA-seq data were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE159288 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159288>]. Source data are provided with this paper. Additional data that support the findings of this study are available from the corresponding authors on reasonable request.

Contributions

P.S.V., E.A.A., E.A.L., M.R.F., R.R.M., M.A.B.M., J.B.L.C., J.A.D. performed enzyme assays. P.S.V. and L.M.Z. performed biophysical analyses. P.S.V., A.G. and M.S.B. performed carbohydrate extraction. P.S.V., E.A.A., M.A.B.M., I.P. and M.T.M. performed

crystallographic studies. S.A.R. synthesized and analyzed the acetylated mono- and disaccharides. A.R.L., I.M.B., C.E.B. and P.O.G. performed gene knockout and in vivo experiments. I.M.B., D.A.A.P., G.F.P. and P.O.G. performed RNA-seq analyses. G.F.P. performed phylogenetic analyses. T.B.L., R.A.S.P., M.S.B. and F.C.G. performed and analyzed mass spectrometry experiments. I.P., P.O.G. and M.T.M. coordinated the work, analyzed the results and wrote the manuscript. P.S.V. and I.M.B. also contributed to the writing of the manuscript.

Acknowledgements

We are grateful to Dr. Felipe Rafael Torres for the discussions on the RT-qPCR analysis, to Dr. Nicolas Terrapon and the CAZy database for analyzing and establishing the novel CE20 family, to Dr. Juliana Velasco de Castro Oliveira for the support in the cultivation of *A. thaliana* and all the staff of the several facilities listed below for their technical support. We thank the Brazilian Synchrotron Light Laboratory (LNLS—CNPEM/MCTI) for the provision of time on the SAXS1 and MX2 beamlines, the Stanford Synchrotron Radiation National Accelerator Laboratory (SSRL-SLAC) for the provision of time on the BL9-2 beamline, the Brazilian Biosciences National Laboratory (LNBio—CNPEM/MCTI) for access to the crystallization (Robolab) facility, the Nuclear Magnetic Resonance (NMR) facility and the Chemistry and Natural Products Laboratory, and the Brazilian Biorenewables National Laboratory (LNBR—CNPEM/MCTI) for the use of the Next Generation Sequencing (NGS) and Characterization of Macromolecules (MAC) open access facilities. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. This research was funded by São Paulo Research Foundation (FAPESP, 2015/26982-0 to M.T.M. and 2015/13684-0 to I.P.) and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 303988-2015-5 to I.P., and 408600/2018-7 to G.F.P.). P.S.V., R.R.M., M.A.B.M., and A.G. received post-doctoral FAPESP fellowships (2016/06509-0, 2017/14253-9, 2016/19995-0 and 2019/13936-0 respectively), I.M.B. and J.A.D. received FAPESP PhD fellowships (2017/00203-0 and 2018/03724-3 respectively), and E.A.A. received CNPq PhD fellowship (158752/2015-5).

References

1. Eckardt, N. A. Role of xyloglucan in primary cell walls. *Plant Cell* **20**, 1421–1422 (2008).
2. O'Neill, M. A. & York, W. S. The composition and structure of plant primary cell walls. in *Annual Plant Reviews Ch. 1* **8**, 1–54 (John Wiley & Sons, Ltd, Chichester, 2018).
3. Waldron, K. W. & Faulds, C. B. Cell wall polysaccharides: composition and structure. in *Comprehensive Glycoscience Ch. 5* **1**, 181–201 (Elsevier, Amsterdam, 2007).
4. Nishinari, K., Takemasa, M., Zhang, H. & Takahashi, R. Storage plant polysaccharides: xyloglucans, galactomannans, glucomannans. in *Comprehensive Glycoscience Ch. 19* **2**, 613–652 (Elsevier, Amsterdam, 2007).
5. Larsbrink, J. *et al.* A complex gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio japonicus*. *Mol. Microbiol.* **94**, 418–433 (2014).
6. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* **506**, 498–502 (2014).
7. Ryan, R. P. *et al.* Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **9**, 344–355 (2011).
8. Timilsina, S. *et al.* *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nat. Rev. Microbiol.* **18**, 415–427 (2020).
9. Brunings, A. M. & Gabriel, D. W. *Xanthomonas citri*: Breaking the surface. *Mol. Plant Pathol.* **4**, 141–157 (2003).
10. Kay, S. & Bonas, U. How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* **12**, 37–43 (2009).
11. Terrapon, N. *et al.* PULDB: The expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* **46**, D677–D683 (2018).
12. Déjean, G., Tauzin, A. S., Bennett, S. W., Creagh, A. L. & Brumer, H. Adaptation of syntenic xyloglucan utilization loci of human gut *Bacteroidetes* to polysaccharide side chain diversity. *Appl. Environ. Microbiol.* **85**, (2019).
13. de Souza, A. P., Leite, D. C. C., Pattathil, S., Hahn, M. G. & Buckeridge, M. S. Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production. *Bioenergy Res.* **6**, 564–579 (2013).
14. Vogel, J. Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11**, 301–307 (2008).
15. Brennan, M., Fakharuzi, D. & Harris, P. J. Occurrence of fucosylated and non-fucosylated xyloglucans in the cell walls of monocotyledons: An immunofluorescence study. *Plant Physiol. Biochem.* **139**, 428–434 (2019).
16. Brennan, M. & Harris, P. J. Distribution of fucosylated xyloglucans among the walls of different cell types in monocotyledons determined by immunofluorescence microscopy. *Mol. Plant* **4**, 144–156 (2011).
17. An, S. Q. *et al.* Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiol. Rev.* **44**, 1–32 (2019).
18. Matsuzawa, T. & Yaoi, K. GH74 xyloglucanases: structures and modes of activity. *Trends Glycosci. Glycotechnol.* **28**, E63–E70 (2016).
19. Arnal, G. *et al.* Substrate specificity, regiospecificity, and processivity in glycoside hydrolase

- family 74. *J. Biol. Chem.* **294**, 13233–13247 (2019).
20. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
 21. Zhong, R., Cui, D. & Ye, Z. H. Xyloglucan *O*-acetyltransferases from *Arabidopsis thaliana* and *Populus trichocarpa* catalyze acetylation of fucosylated galactose residues on xyloglucan side chains. *Planta* **248**, 1159–1171 (2018).
 22. Nardini, M. & Dijkstra, B. W. α/β hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.* **9**, 732–737 (1999).
 23. Montanier, C. *et al.* The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol.* **7**, 0687–0697 (2009).
 24. Correia, M. A. S. *et al.* Crystal structure of a cellulosomal family 3 carbohydrate esterase from *Clostridium thermocellum* provides insights into the mechanism of substrate recognition. *J. Mol. Biol.* **379**, 64–72 (2008).
 25. Bitto, E. *et al.* The structure at 1.6 Å resolution of the protein product of the At4g34215 gene from *Arabidopsis thaliana*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1655–1661 (2005).
 26. Mølgaard, A., Kauppinen, S. & Larsen, S. Rhamnogalacturonan acetyltransferase elucidates the structure and function of a new family of hydrolases. *Structure* **8**, 373–383 (2000).
 27. Michalak, L. *et al.* A pair of esterases from a commensal gut bacterium remove acetylations from all positions on complex β -mannans. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7122–7130 (2020).
 28. Sun, Y. *et al.* Molecular basis of the general base catalysis of an α/β -hydrolase catalytic triad. *J. Biol. Chem.* **289**, 15867–15879 (2014).
 29. Rauwerdink, A. & Kazlauskas, R. J. How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of α/β -hydrolase fold enzymes. *ACS Catal.* **5**, 6153–6176 (2015).
 30. Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H. & Shaw, J. F. GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534–552 (2004).
 31. Ménard, R. & Storer, A. C. Oxyanion hole interactions in serine and cysteine proteases. *Biol. Chem. Hoppe. Seyler.* **373**, 393–400 (1992).
 32. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
 33. Nagae, M. *et al.* Structural basis of the catalytic reaction mechanism of novel 1,2- α -L-fucosidase from *Bifidobacterium bifidum*. *J. Biol. Chem.* **282**, 18497–18509 (2007).
 34. Sela, D. A. *et al.* *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α -fucosidases are active on fucosylated human milk oligosaccharides. *Appl. Environ. Microbiol.* **78**, 795–803 (2012).
 35. Yang, T. C., Hu, R. M., Weng, S. F. & Tseng, Y. H. Identification of a hypothetical protein of plant pathogenic *Xanthomonas campestris* as a novel β -galactosidase. *J. Mol. Microbiol. Biotechnol.* **13**, 172–180 (2007).
 36. Cao, H., Walton, J. D., Brumm, P. & Phillips, G. N. Crystal structure of α -xylosidase from *Aspergillus niger* in complex with a hydrolyzed xyloglucan product and new insights in accurately predicting substrate specificities of GH31 family glycosidases. *ACS Sustain. Chem. Eng.* **8**, 2540–2547 (2020).
 37. Larsbrink, J. *et al.* Structural and enzymatic characterization of a glycoside hydrolase family 31

- α -xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochem. J.* **436**, 567–580 (2011).
38. Rigden, D. J., Mello, L. V. & Galperin, M. Y. The PA14 domain, a conserved all- β domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem. Sci.* **29**, 335–339 (2004).
 39. Yoshida, E. *et al.* Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β -glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39–49 (2010).
 40. Hemsworth, G. R. *et al.* Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut. *Open Biol.* **6**, 160142 (2016).
 41. Silipo, A. *et al.* NMR spectroscopic analysis reveals extensive binding interactions of complex xyloglucan oligosaccharides with the *Cellvibrio japonicus* glycoside hydrolase family 31 α -xylosidase. *Chem. - A Eur. J.* **18**, 13395–13404 (2012).
 42. Nelson, C. E. *et al.* Comprehensive functional characterization of the glycoside hydrolase family 3 enzymes from *Cellvibrio japonicus* reveals unique metabolic roles in biomass saccharification. *Environ. Microbiol.* **19**, 5025–5039 (2017).
 43. Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* **82**, 1686–1692 (2016).
 44. Hsieh, Y. S. Y. & Harris, P. J. Xyloglucans of monocotyledons have diverse structures. *Mol. Plant* **2**, 943–965 (2009).
 45. Hoffman, M. *et al.* Structural analysis of xyloglucans in the primary cell walls of plants in the subclass *Asteridae*. *Carbohydr. Res.* **340**, 1826–1840 (2005).
 46. Facincani, A. P. *et al.* Comparative proteomic analysis reveals that T3SS, Tfp, and xanthan gum are key factors in initial stages of *Citrus sinensis* infection by *Xanthomonas citri* subsp. *citri*. *Funct. Integr. Genomics* **14**, 205–217 (2014).
 47. Guo, Y., Figueiredo, F., Jones, J. & Wang, N. HrpG and HrpX play global roles in coordinating different virulence traits of *Xanthomonas axonopodis* pv. *citri*. *Mol. Plant-Microbe Interact.* **24**, 649–661 (2011).
 48. de Melo, R. R. *et al.* Identification of a cold-adapted and metal-stimulated β -1,4-glucanase with potential use in the extraction of bioactive compounds from plants. *Int. J. Biol. Macromol.* **166**, 190–199 (2021).
 49. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henrissat, B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12**, 186 (2012).
 50. Hudson, K. L. *et al.* Carbohydrate-aromatic interactions in proteins. *J. Am. Chem. Soc.* **137**, 15152–15160 (2015).
 51. Arlat, M., Gough, C. L., Barber, C. E., Boucher, C. & Daniels, M. J. *Xanthomonas campestris* contains a cluster of *hrp* genes related to the larger *hrp* cluster of *Pseudomonas solanacearum*. *Mol. Plant. Microbe. Interact.* **4**, 593–601 (1991).
 52. Wengelnik, K., Marie, C., Russel, M. & Bonas, U. Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction. *J. Bacteriol.* **178**, 1061–1069 (1996).
 53. Tsuge, S. *et al.* Expression of *Xanthomonas oryzae* pv. *oryzae* *hrp* genes in XOM2, a novel synthetic medium. *J. Gen. Plant Pathol.* **68**, 363–371 (2002).

54. Ikawa, Y. & Tsuge, S. The quantitative regulation of the *hrp* regulator HrpX is involved in sugar-source-dependent *hrp* gene expression in *Xanthomonas oryzae* pv. *oryzae*. *FEMS Microbiol. Lett.* **363**, fnw071 (2016).
55. Ikawa, Y., Ohnishi, S., Shoji, A., Furutani, A. & Tsuge, S. Concomitant regulation by a LacI-Type transcriptional repressor XylR on genes involved in xylan and xylose metabolism and the type III secretion system in rice pathogen *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant-Microbe Interact.* **31**, 605–613 (2018).
56. Rashid, M. M., Ikawa, Y. & Tsuge, S. GamR, the LysR-type galactose metabolism regulator, regulates *hrp* gene expression via transcriptional activation of two key *hrp* regulators, HrpG and HrpX, in *Xanthomonas oryzae* pv. *oryzae*. *Appl. Environ. Microbiol.* **82**, 3947–3958 (2016).
57. Déjean, G. *et al.* The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv. *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytol.* **198**, 899–915 (2013).
58. Santos, C. R. *et al.* Molecular mechanisms associated with xylan degradation by *Xanthomonas* plant pathogens. *J. Biol. Chem.* **289**, 32186–32200 (2014).
59. Tayi, L., Maku, R. V., Patel, H. K. & Sonti, R. V. Identification of pectin degrading enzymes secreted by *Xanthomonas oryzae* pv. *oryzae* and determination of their role in virulence on rice. *PLoS One* **11**, 1–15 (2016).
60. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
61. Na, S. I. *et al.* UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol.* **56**, 281–285 (2018).
62. de Sousa, A. S. *et al.* A rationally identified marine GH1 β -glucosidase has distinguishing functional features for simultaneous saccharification and fermentation. *Biofuels, Bioprod. Biorefining* **14**, 1163–1179 (2020).
63. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
64. Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F. & Polikarpov, I. SAXSMoW 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci.* **28**, 454–463 (2019).
65. Krissinel, E. & Henrick, K. ‘Protein interfaces, surfaces and assemblies’ service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html), inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
66. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
67. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
68. Martinez-Fleites, C. *et al.* Crystal structures of *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. *J. Biol. Chem.* **281**, 24922–24933 (2006).
69. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485 (2010).
70. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).
71. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta*

- Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
72. Davis, I. W. *et al.* MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
 73. Zheng, H. *et al.* Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat. Protoc.* **9**, 156–170 (2014).
 74. Zheng, H. *et al.* CheckMyMetal: A macromolecular metal-binding validation tool. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 223–233 (2017).
 75. Handing, K. B. *et al.* Characterizing metal-binding sites in proteins with X-ray crystallography. *Nat. Protoc.* **13**, 1062–1090 (2018).
 76. Agirre, J. *et al.* Privateer: Software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).
 77. Buckeridge, M. S., Rocha, D. C., Reid, J. S. G. & Dietrich, S. M. C. Xyloglucan structure and post-germinative metabolism in seeds of *Copaifera langsdorfii* from savanna and forest populations. *Physiol. Plant.* **86**, 145–151 (1992).
 78. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497 (1962).
 79. Lerouxel, O. *et al.* Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting. *Plant Physiol.* **130**, 1754–1763 (2002).
 80. Höfle, G., Steglich, W. & Vorbrüggen, H. 4-dialkylaminopyridines as highly active acylation catalysts. [new synthetic method (25)]. *Angew. Chemie Int. Ed. English* **17**, 569–583 (1978).
 81. Dos Santos, C. R. *et al.* The mechanism by which a distinguishing arabinofuranosidase can cope with internal di-substitutions in arabinoxylans. *Biotechnol. Biofuels* **11**, 223 (2018).
 82. Chomczynski, P. & Mackey, K. Modification of the TRI Reagent(TM) procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *Biotechniques* **19**, 942–945 (1995).
 83. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
 84. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
 85. Da Silva, A. C. R. *et al.* Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463 (2002).
 86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 87. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
 88. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299 (1996).
 89. Carmona, R. *et al.* Automated identification of reference genes based on RNA-seq data. *Biomed. Eng. Online* **16**, 65 (2017).
 90. Hoang, V. L. T. *et al.* RNA-seq reveals more consistent reference genes for gene expression studies in human non-melanoma skin cancers. *PeerJ* **2017**, e3631 (2017).
 91. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper -

- Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
92. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
 93. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
 94. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45e – 45 (2001).
 95. Ried, J. L. & Collmer, A. An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* **57**, 239–246 (1987).
 96. Soprano, A. S. *et al.* Crystal Structure and Regulation of the Citrus Pol III Repressor MAF1 by Auxin and Phosphorylation. *Structure* **25**, 1360-1370.e4 (2017).

Reprint Permission

- This article can be cited as:

Vieira, P.S., Bonfim, I.M., Araujo, E.A. *et al.* Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat Commun* **12**, 4049 (2021). <https://doi.org/10.1038/s41467-021-24277-4>



RightsLink



Help ▾



Live Chat

Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors

SPRINGER NATURE

Author: Plinio S. Vieira et al

Publication: Nature Communications

Publisher: Springer Nature

Date: Jun 30, 2021

Copyright © 2021, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact [Springer Nature](#)

- Accession:

<https://s100.copyright.com/AppDispatchServlet?title=Xyloglucan%20processing%20machinery%20in%20Xanthomonas%20pathogens%20and%20its%20role%20in%20the%20transcriptional%20activation%20of%20virulence%20factors&author=Plinio%20S.%20Vieira%20et%20al&contentID=10.1038%2Fs41467-021-24277-4©right=The%20Author%28s%29&publication=2041-1723&publicationDate=2021-06-30&publisherName=SpringerNature&orderBeanReset=true&oa=CC%20BY>

Supplementary Information

Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors

Plinio S. Vieira^{1#}, Isabela M. Bonfim^{1,2#}, Evandro A. Araujo^{1,3}, Ricardo R. Melo¹, Augusto R. Lima¹, Melissa R. Fessel⁴, Douglas A. A. Paixão¹, Gabriela F. Persinoti¹, Silvana A. Rocco⁵, Tatiani B. Lima¹, Renan A. S. Pirolla¹, Mariana A. B. Morais¹, Jessica B. L. Correa¹, Leticia M. Zanphorlin¹, Jose A. Diogo^{1,2}, Evandro A. Lima¹, Adriana Grandis⁶, Marcos S. Buckeridge⁶, Fabio C. Gozzo⁷, Celso E. Benedetti⁵, Igor Polikarpov⁸, Priscila O. Giuseppe^{1*} and Mario T. Murakami^{1*}

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil.

³Brazilian Synchrotron Light Laboratory (LNLS), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

⁴Butantan Institute, Butantan Foundation, São Paulo, São Paulo, Brazil.

⁵Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

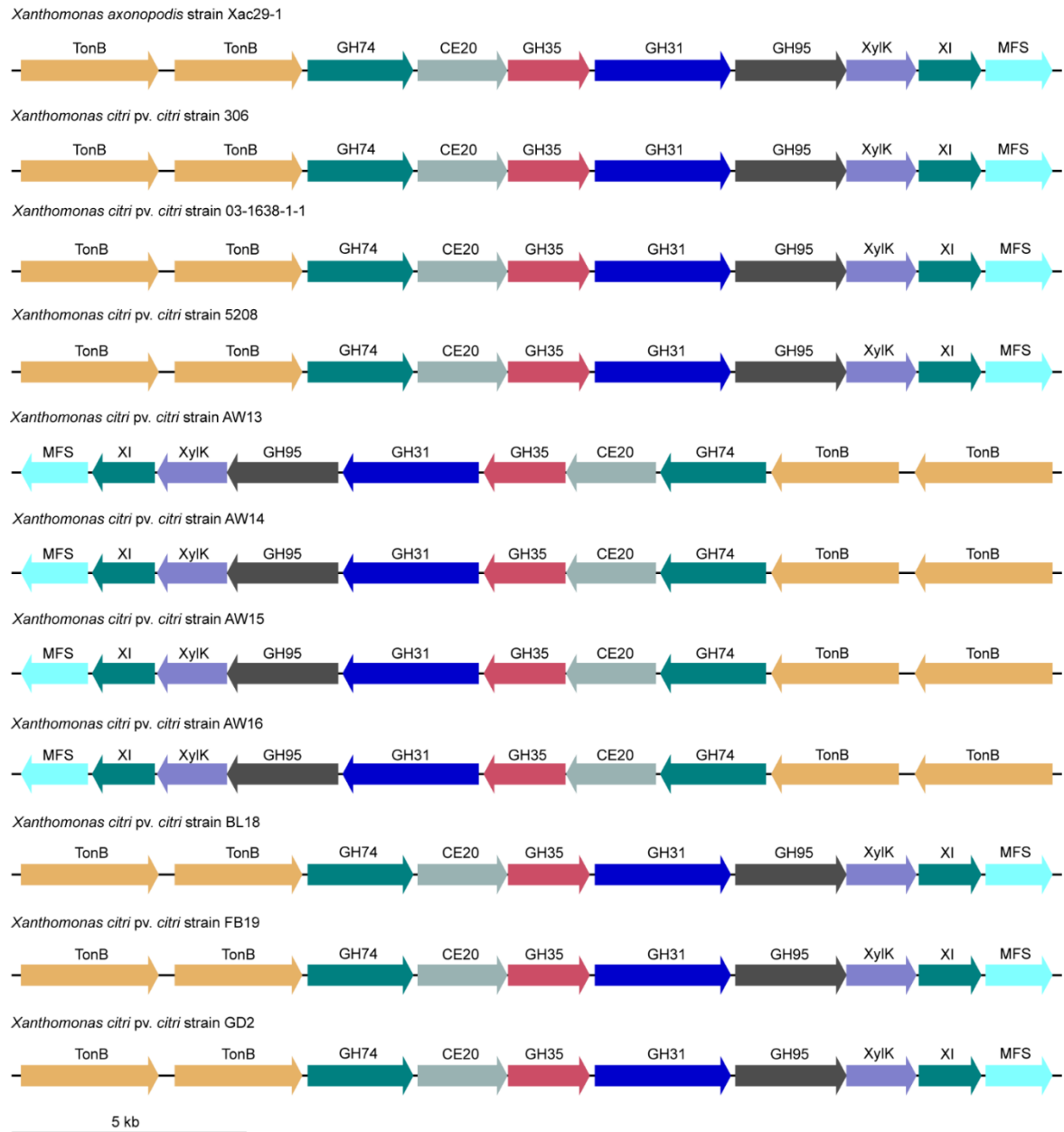
⁶Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

⁷Institute of Chemistry, University of Campinas, Campinas, São Paulo, Brazil.

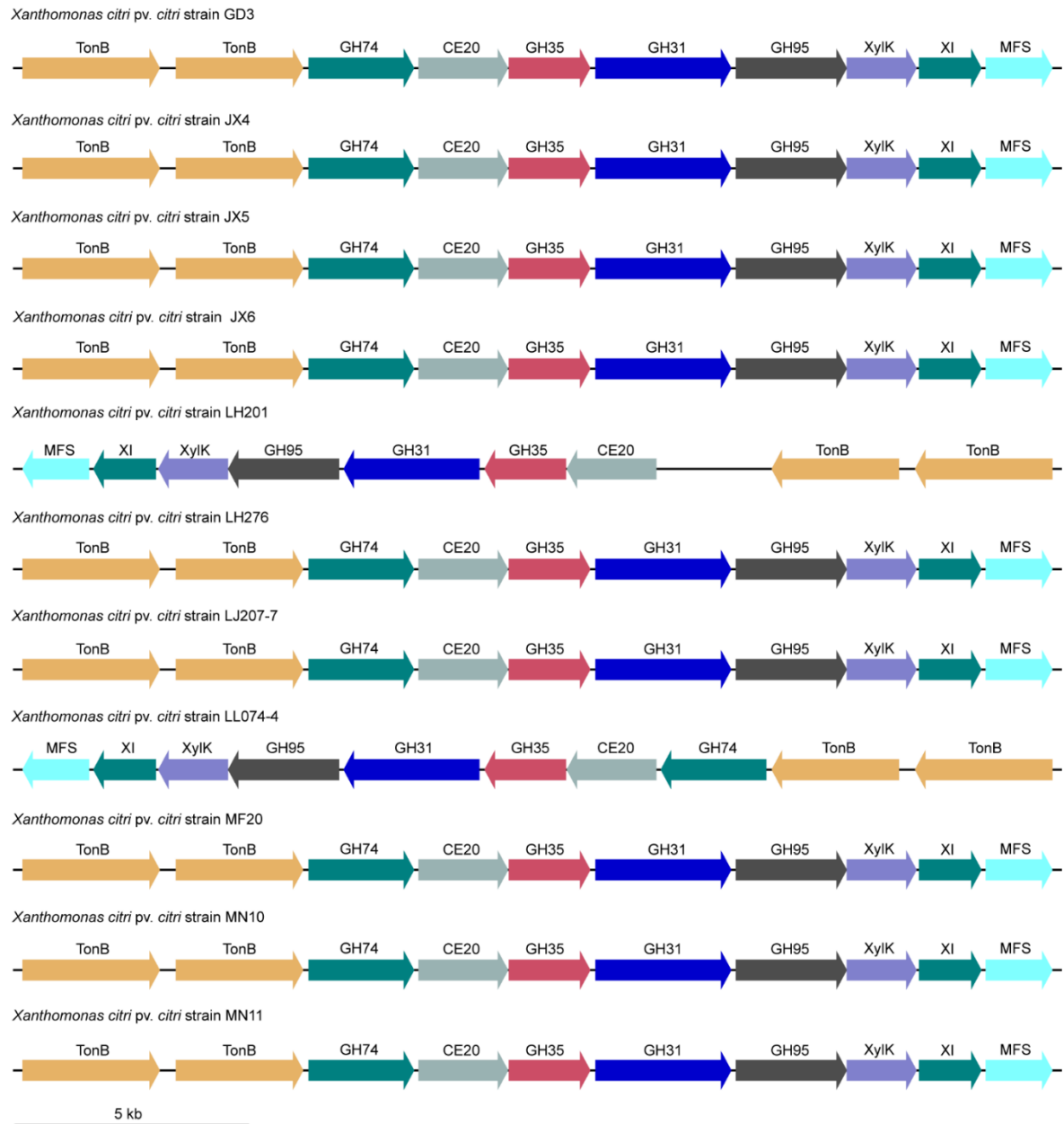
⁸São Carlos Institute of Physics, University of São Paulo, São Carlos, São Paulo, Brazil.

[#]These authors have equally contributed to this work

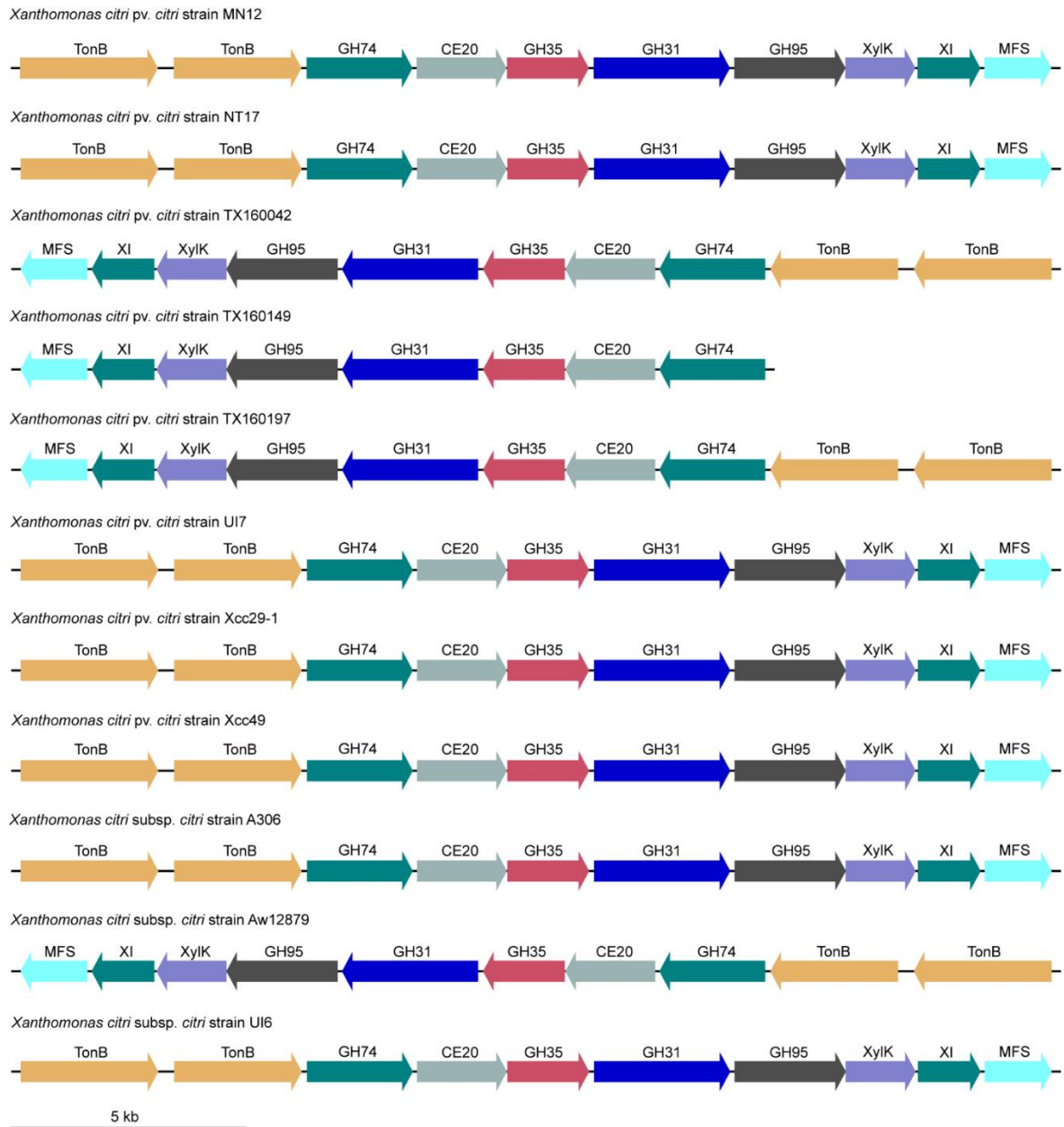
*Correspondence and requests should be addressed to M.T.M (mario.murakami@lnbr.cnpem.br) or P.O.G (priscila.giuseppe@lnbr.cnpem.br)



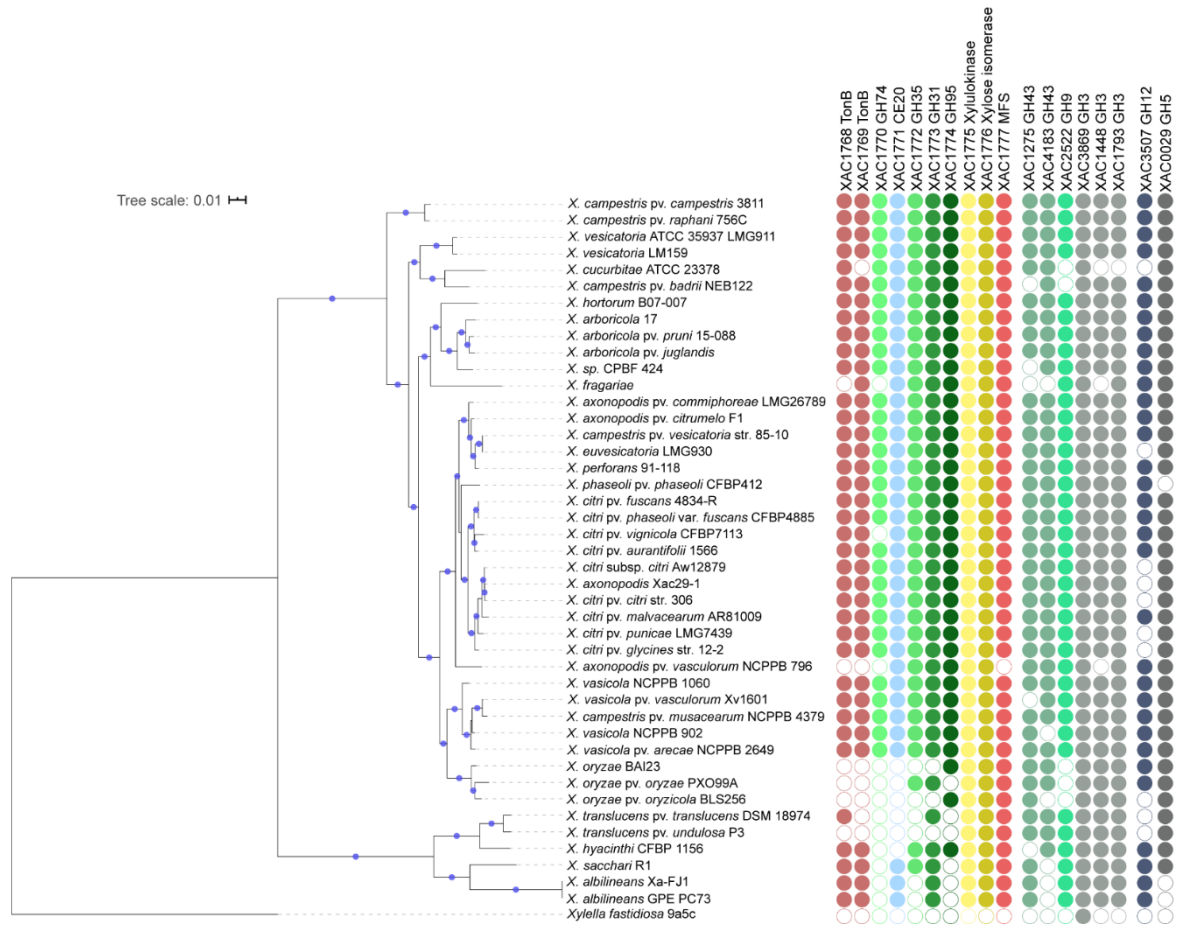
Supplementary Figure 1. XyGUL genes are conserved among *Xanthomonas citri* strains. Conservation of XyGUL genes and downstream genes corresponding to xylose metabolism and MFS transporter was evaluated in all complete genomes of *X. citri* pv. *citri* currently available at the Refseq database. Each gene was colored according to the family it belongs to or its functional annotation.

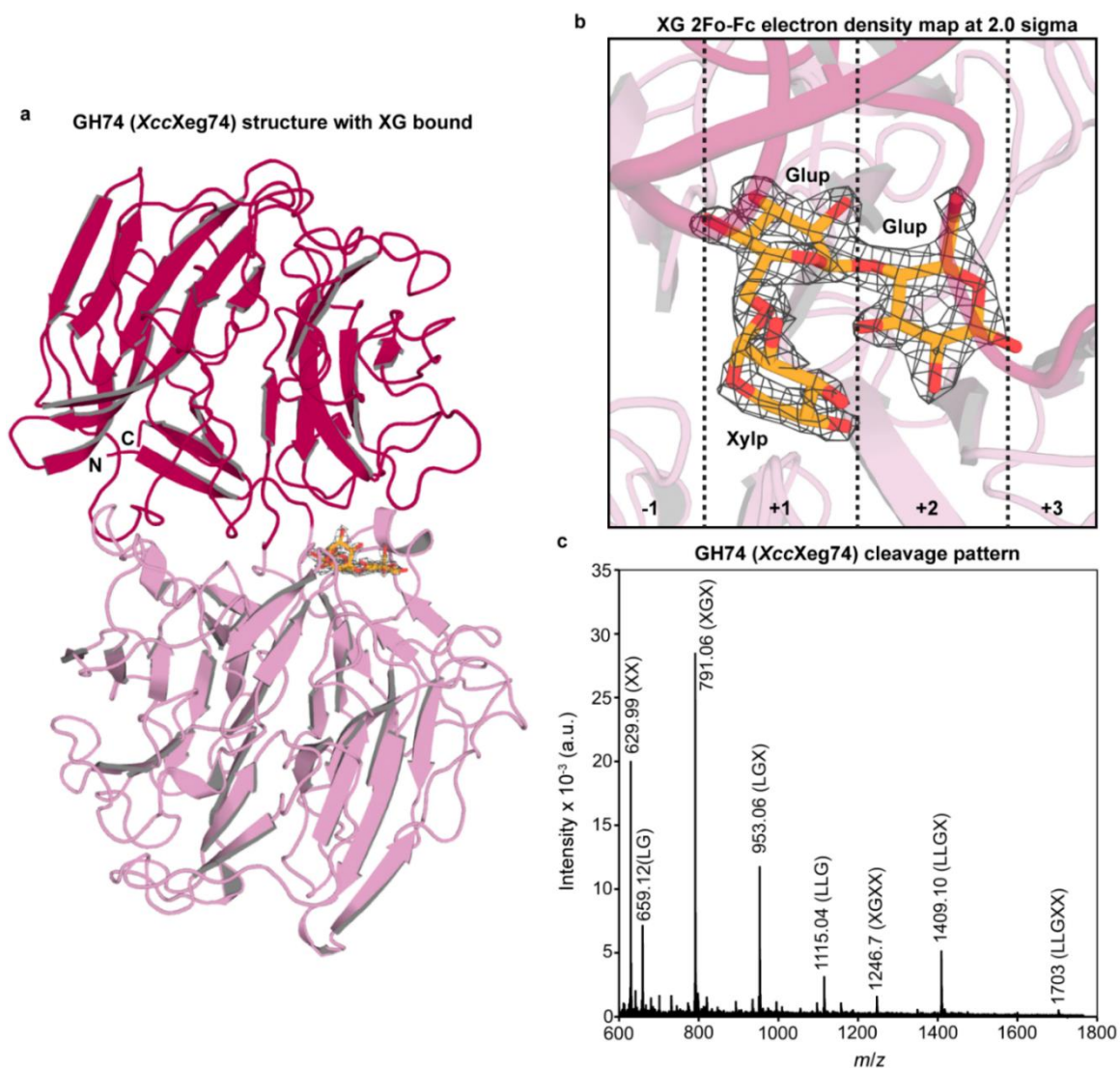


Supplementary Figure 1. (Continued) XyGUL genes are conserved among *Xanthomonas citri* strains. Conservation of XyGUL genes and downstream genes corresponding to xylose metabolism and MFS transporter was evaluated in all complete genomes of *X. citri* pv. *citri* currently available at the Refseq database. Each gene was colored according to the family it belongs to or its functional annotation.

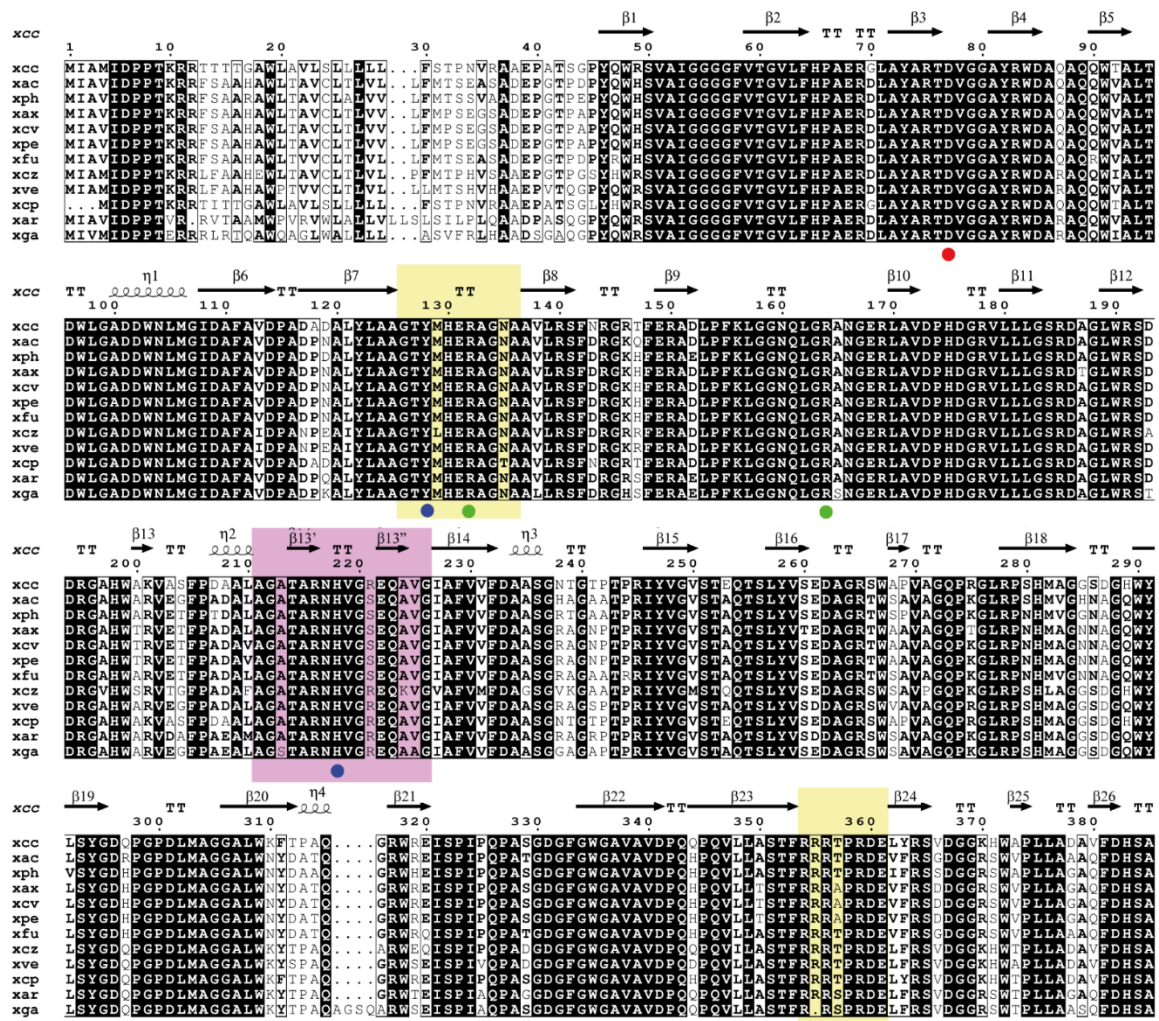


Supplementary Figure 1. (Continued) XyGUL genes are conserved among *Xanthomonas citri* strains. Conservation of XyGUL genes and downstream genes corresponding to xylose metabolism and MFS transporter was evaluated in all complete genomes of *X. citri* pv. *citri* currently available at the Refseq database. Each gene was colored according to the family it belongs to or its functional annotation.

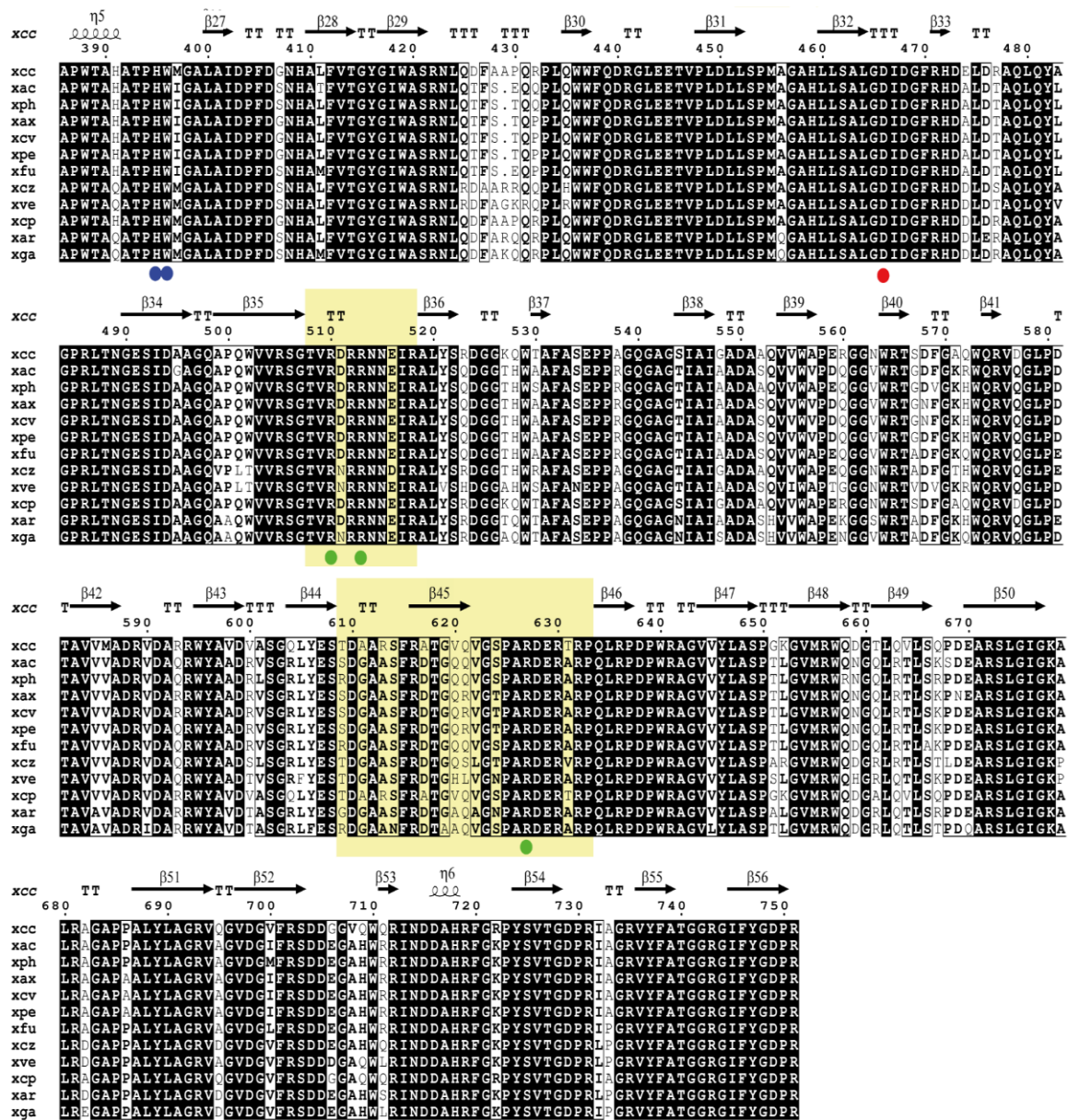




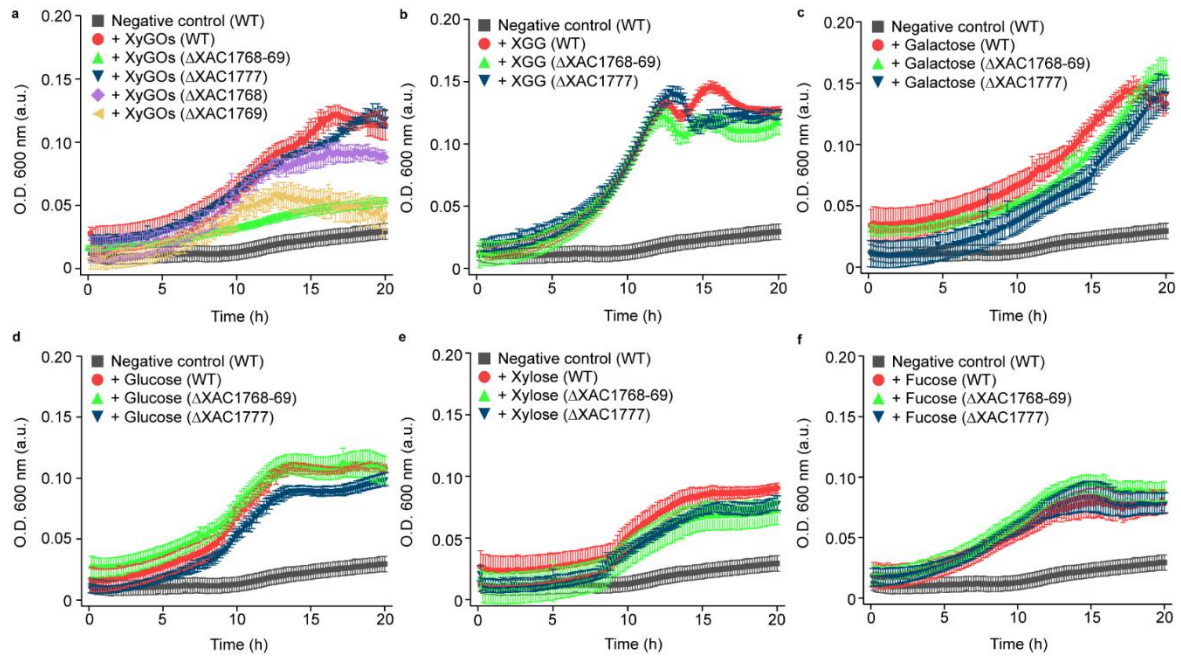
Supplementary Figure 3. Structure of *Xanthomonas* GH74 (*XccXeg74*) in complex with XG and mass spectrometry spectrum of cleavage products. (a) Cartoon representation of the crystal structure of *XccXeg74* monomer with XG oligosaccharide (sticks, yellow C-atoms) bound to the positive subsite region. The two seven-bladed β -propellers are shown in shades of pink. XG = α -D-xylopyranosyl-(1->6)- β -D-glucopyranosyl-(1->4)- β -D-glucopyranose. (b) Structural details of the active site including the XG oligosaccharide, and the 2F_o-F_c electron density map contoured at 2 σ level. Numbers -1, +1, +2, +3 indicate subsites. Glup = β -D-glucopyranosyl moiety. Xylp = α -D-xylopyranosyl moiety. (c) MALDI-TOF spectrum of the GH74 released products from *Copaifera langsdorfii* xyloglucan. Numbers above the peaks represent *m/z* values of each product (see Fig. 2). Letters indicate the type of substitutions appended to the glucose backbone of the identified oligosaccharides. G = non-substituted glucose; X = glucose substituted with a xylose at C-6; L = X with a galactose appended at xylose C-2.



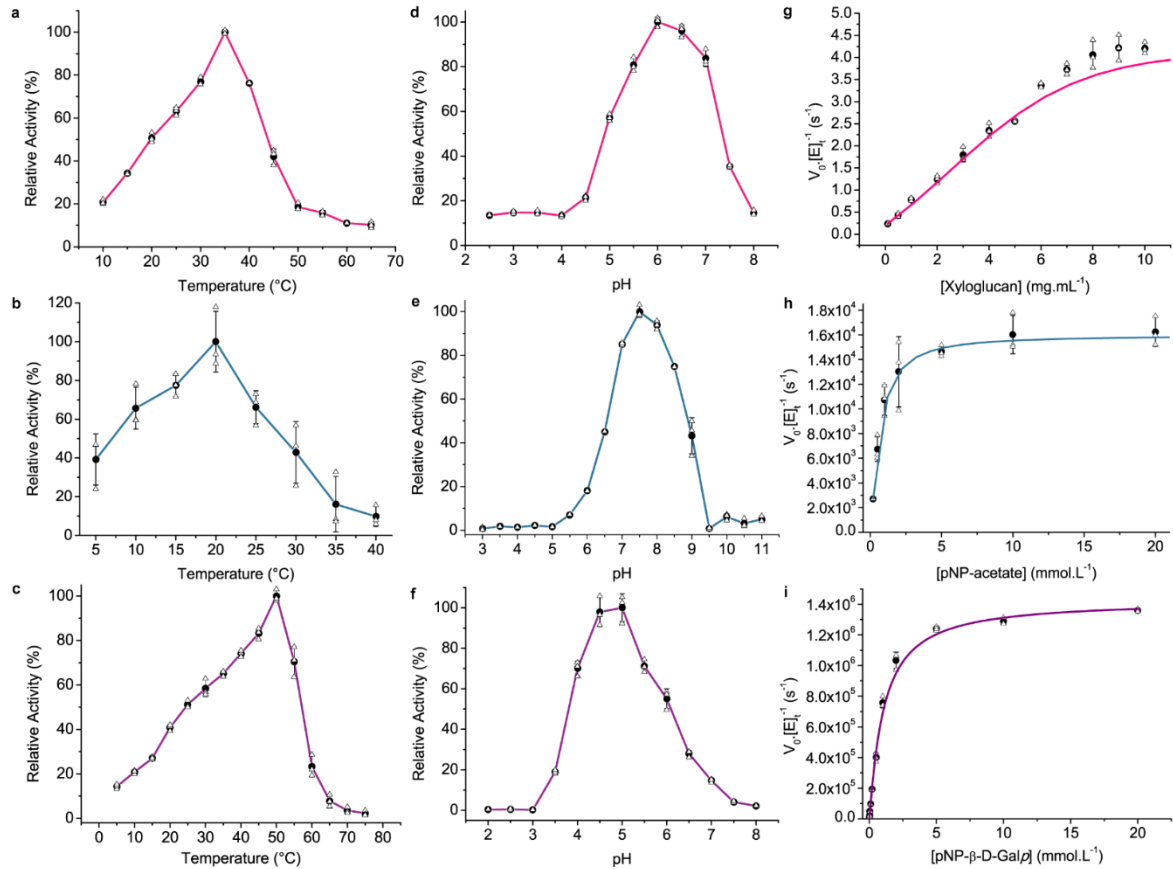
Supplementary Figure 4. Alignment of GH74 sequences from representative *Xanthomonas* species. Numbers above the alignment represent the amino acid residues numbers of *XccXeg74* in the sequence alignment. Secondary structure elements are also depicted above and labeled according to the crystallographic model of *XccXeg74*. Arrows = β -sheets, helices = α -helices and T = β -turn. Yellow boxes delimit the loop extensions observed in *Xanthomonas* enzymes. The light purple box indicates the β -hairpin insertion conserved across *Xanthomonas* sequences. Green circles mark the arginine residues involved in substrate binding, blue circles indicate the residues that make stacking interactions with the substrate, and red circles identify the catalytic residues D77 (acid) and D466 (base) (See Fig. 2). Residues in black boxes are fully conserved. The representative species are *xcc*: *Xanthomonas campestris* pv. *campestris* ATCC 33913, *xac*: *Xanthomonas citri* pv. *citri* 306, *xph*: *Xanthomonas phaseoli* pv. *phaseoli* CFBP6546R, *xax*, *Xanthomonas axonopodis* pv. *citrumelo* F1, *xcv*, *Xanthomonas campestris* pv. *vesicatoria* 85-10, *xpe*, *Xanthomonas perforans* LH3, *xfu*, *Xanthomonas citri* pv. *fuscans* 4834-R, *xcz*, *Xanthomonas cucurbitae* ATCC 23378, *xve*, *Xanthomonas vesicatoria* ATCC 35937 LMG911, *xcp*, *Xanthomonas campestris* pv. *raphani* 756C, *xar*, *Xanthomonas arboricola* 17, *xga*, *Xanthomonas gardneri* ICMP 7383. Alignment performed with the Clustal Ω^1 server and image produced with ESPrpt 3.0 server².



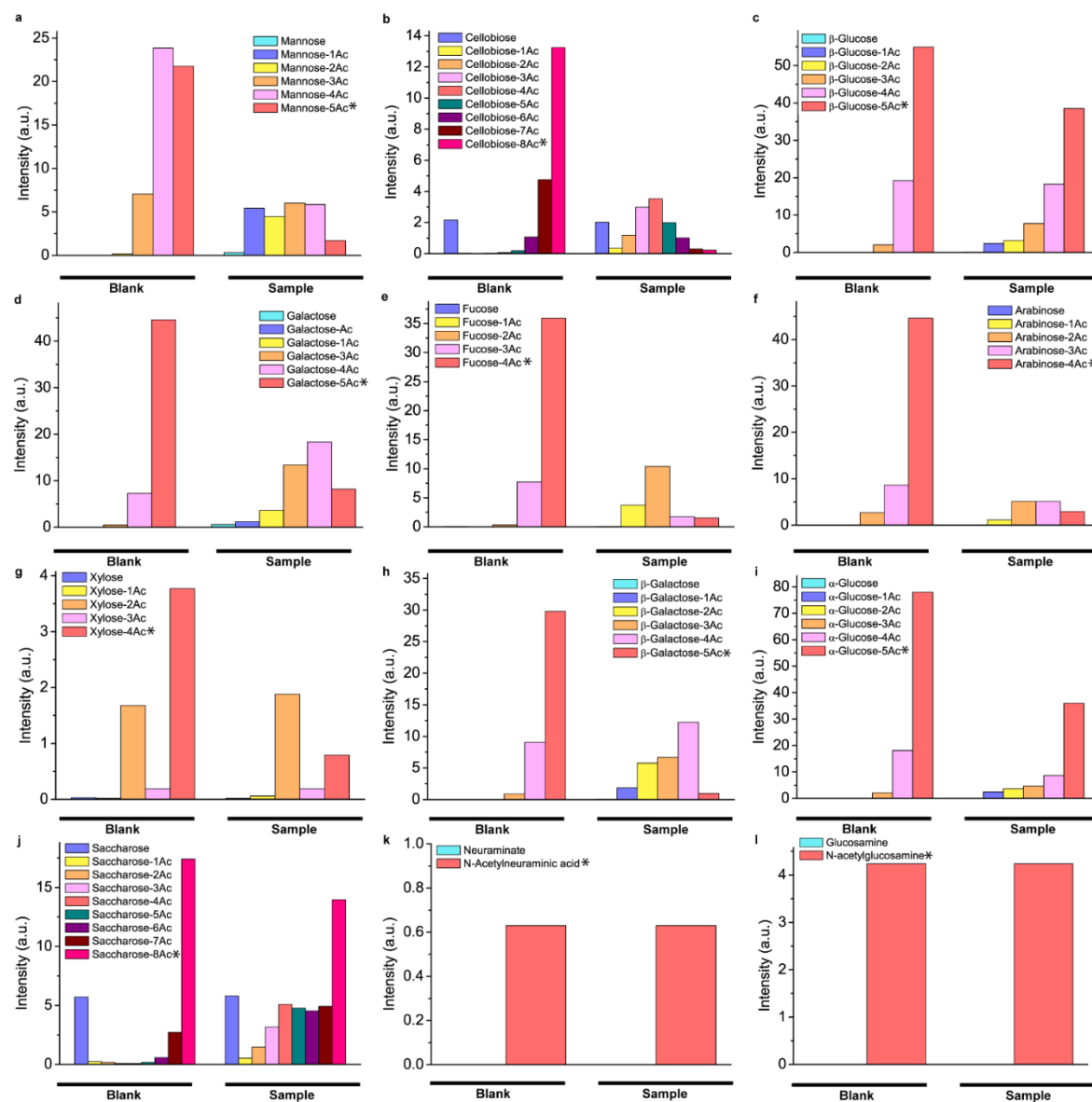
Supplementary Figure 4. Alignment of GH74 sequences from representative *Xanthomonas* species (Continued). Numbers above the alignment represent the amino acid residues numbers of *XccXeg74* in the sequence alignment. Secondary structure elements are also depicted above and labeled according to the crystallographic model of *XccXeg74*. Arrows = β -sheets, helices = α -helices and T = β -turn. Yellow boxes delimit the loops extensions observed for *Xanthomonas* enzymes. The light purple box marks the β -hairpin insertion, also conserved in other *Xanthomonas*. Below the alignment, green circles mark the arginine residues involved in substrate binding, blue circles mark the residues that make stacking interactions with the substrate, and red circles identify the catalytic residues D77 (acid) and D466 (base) (See Fig. 2). Residues in black boxes are fully conserved. The representative species are *xcc*: *Xanthomonas campestris* pv. *campestris* ATCC 33913, *xac*: *Xanthomonas citri* pv. *citri* 306, *xph*: *Xanthomonas phaseoli* pv. *phaseoli* CFBP6546R, *xax*, *Xanthomonas axonopodis* pv. *citrumelo* F1, *xcv*, *Xanthomonas campestris* pv. *vesicatoria* 85-10, *xpe*, *Xanthomonas perforans* LH3, *xfu*, *Xanthomonas citri* pv. *fuscans* 4834-R, *xcz*, *Xanthomonas cucurbitae* ATCC 23378, *xve*, *Xanthomonas vesicatoria* ATCC 35937 LMG911, *xcp*, *Xanthomonas campestris* pv. *raphani* 756C, *xar*, *Xanthomonas arboricola* 17, *xga*, *Xanthomonas gardneri* ICMP 7383. Alignment performed with the Clustal Ω^1 server and image produced by ESPript 3.0 server².



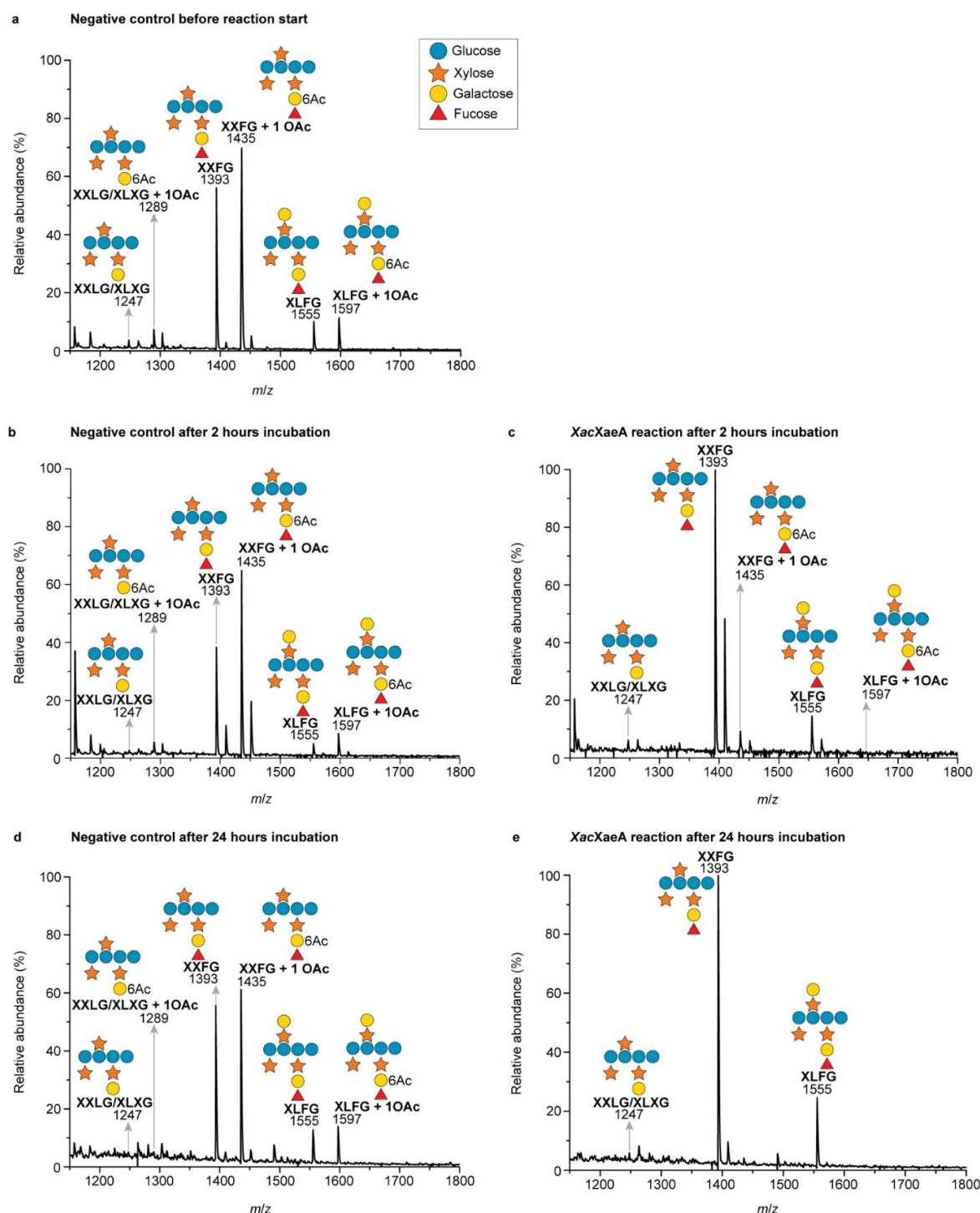
Supplementary Figure 5. Deletion of XyGUL TonB-dependent transporter *cirA* (XAC1769) impaired XyGOs uptake, but the lack of XyGUL MFS transporter did not affect growth. Growth curves of *X. citri* strains (WT in red, Δ XAC1768-69 in green, Δ XAC1768 in lilac, Δ XAC1769 in wheat, Δ XAC1777 in blue) in minimal medium supplemented with XyGOs (a); a mixture of xylose, glucose, and galactose (XGG) in the same molar proportion found in tamarind XyGOs (b); galactose (c); glucose (d); xylose (e); and fucose (f). The OD_{600nm} of WT in minimal medium without carbohydrate was used as a negative control (gray). Data are presented as mean \pm SD from three independent experiments (n=3). Data points are shown as filled symbols as described in each legend. Source data are provided as a source datafile.



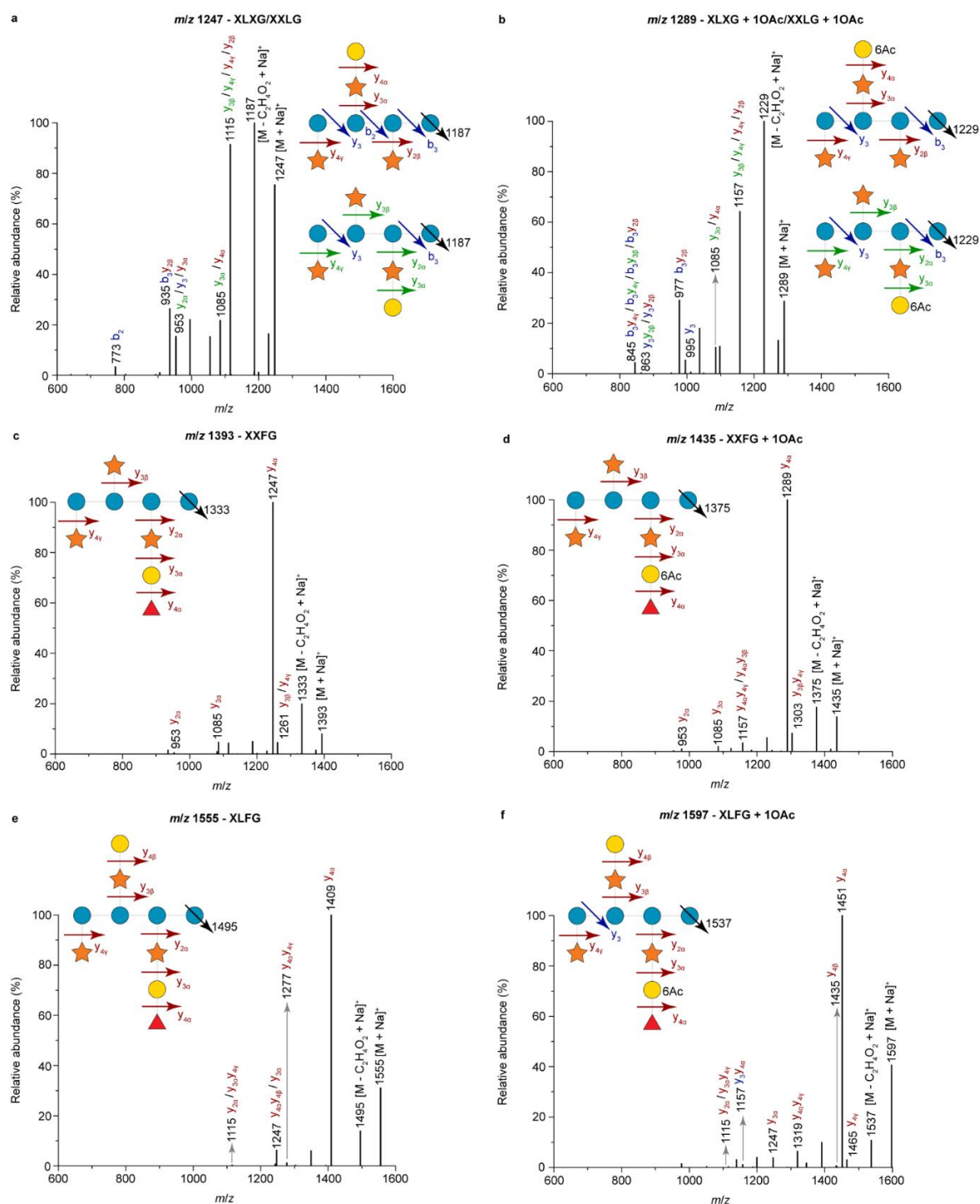
Supplementary Figure 6. Biochemical and kinetic characterization of XyGUL enzymes. *Left*, temperature dependence curves of *XacXeg74* (a), *XacXaeA* (b) and *XacGalD* (c). *Middle*, pH dependence curves of *XacXeg74* (d), *XacXaeA* (e) and *XacGalD* (f). *Right*, substrate saturation curves of *XacXeg74* using tamarind xyloglucan (g), *XacXaeA* using *para*-nitrophenyl-acetate (h) and *XacGalD* using *para*-nitrophenyl- β -D-galactopyranoside (i). Data are shown as mean \pm SD from three independent experiments ($n=3$). (V_0 =initial velocity; $[E]_t$ =enzyme concentration). Mean values are shown as filled circles, while measured data are shown as empty circles. Standard deviations are shown as black lines. Source data are provided as a source data file.



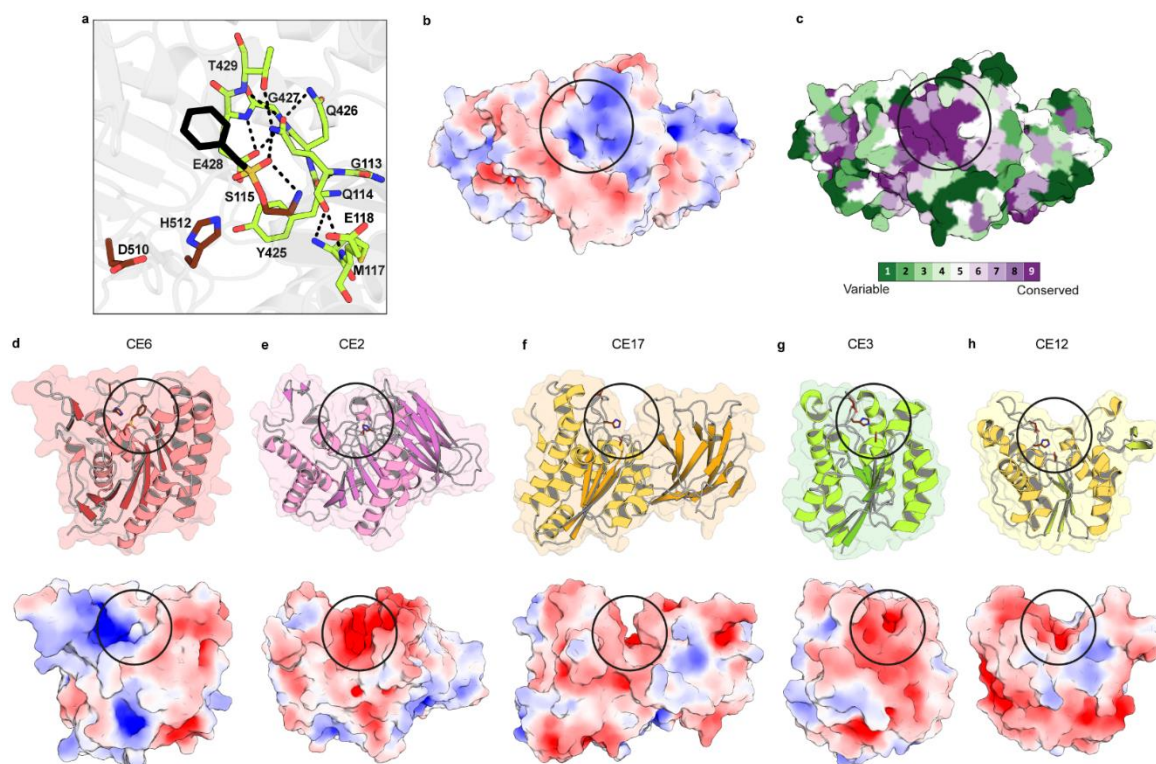
Supplementary Figure 7. Substrate preference of *XacXaeA*. Fully *O*-acetylated mono- or disaccharides were synthesized and used as a library screening for the enzyme activity detection test. *N*- and *O*-acetylated substrates were also used for the bond type preference analysis. Reactions were carried out at 20° C, 80 mmol.L⁻¹ HEPES buffer (pH 7.5) and 600 rpm for 15 min. Samples and blanks (the same respective mix without the enzyme) were collected and added by 40 μ L methanol to stop the reaction. A total of 15 μ L of the quenched reactions were added to 183 μ L of water and 2 μ L of xylo-tetraose (an internal standard used to increase the reliability of the method) and analyzed by ESI (+) mass spectrometry in scan mode (m/z 150-900) (See Methods for details). Each bar represents the detection of acetylated/non-acetylated carbohydrate species as specified in each legend (the symbol * represents the initial substrate used in each reaction). Ac = acetate.



Supplementary Figure 8. *XacXaeA* activity assay on xyloglucan oligosaccharides. Fucosylated polysaccharides were extracted from *A. thaliana* cell walls and cleaved to oligosaccharides by the action of *XacXeg74*. Products were analyzed by ESI(+)-MS/MS and peaks assigned to acetylated or non-acetylated (fuco)galactoxyloglucooligosaccharides considering the literature³⁻⁵. Negative control before reaction (a), after 2 h (b) and 24 h (d) incubation. *XacXaeA* reaction after 2 h (c) and 24 h (e) incubation. Note that most of the acetyl moieties were removed in 2 h, indicating that *XacXaeA* is active on the three identified acetylated oligosaccharides (See Supplementary Table 9 and Supplementary Fig. 9). Carbohydrates are represented by geometric shapes (glucose: blue circles, xylose: orange stars, galactose: yellow circles and fucose: red triangles).



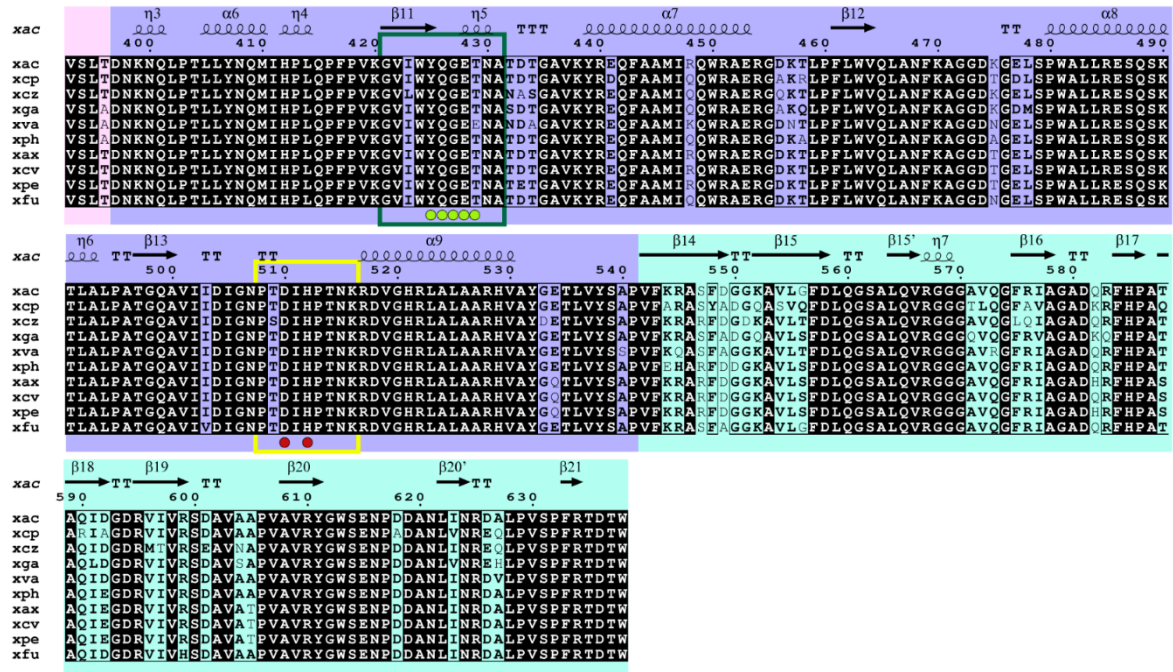
Supplementary Figure 9. Fragmentation patterns of the annotated xyloglucan oligosaccharides assessed by tandem mass spectrometry. In the represented oligosaccharides, arrows indicate the fragmentation position with each ion identified in the corresponding spectra of XLXG/XXLG (a), XXFG (c), XLFG (e) and the acetylated equivalents (b, d, and f, respectively). Internal fragmentations are in black, fragmentations in the main chain are in blue, fragmentations in the side-chains are in red or green to differentiate ions from distinct isomers. See also Supplementary Table 9 and Supplementary Fig. 8. Fragmentations were annotated following the literature⁶. Carbohydrates are represented by geometric shapes (glucose: blue circles, xylose: orange stars, galactose: yellow circles and fucose: red triangles).



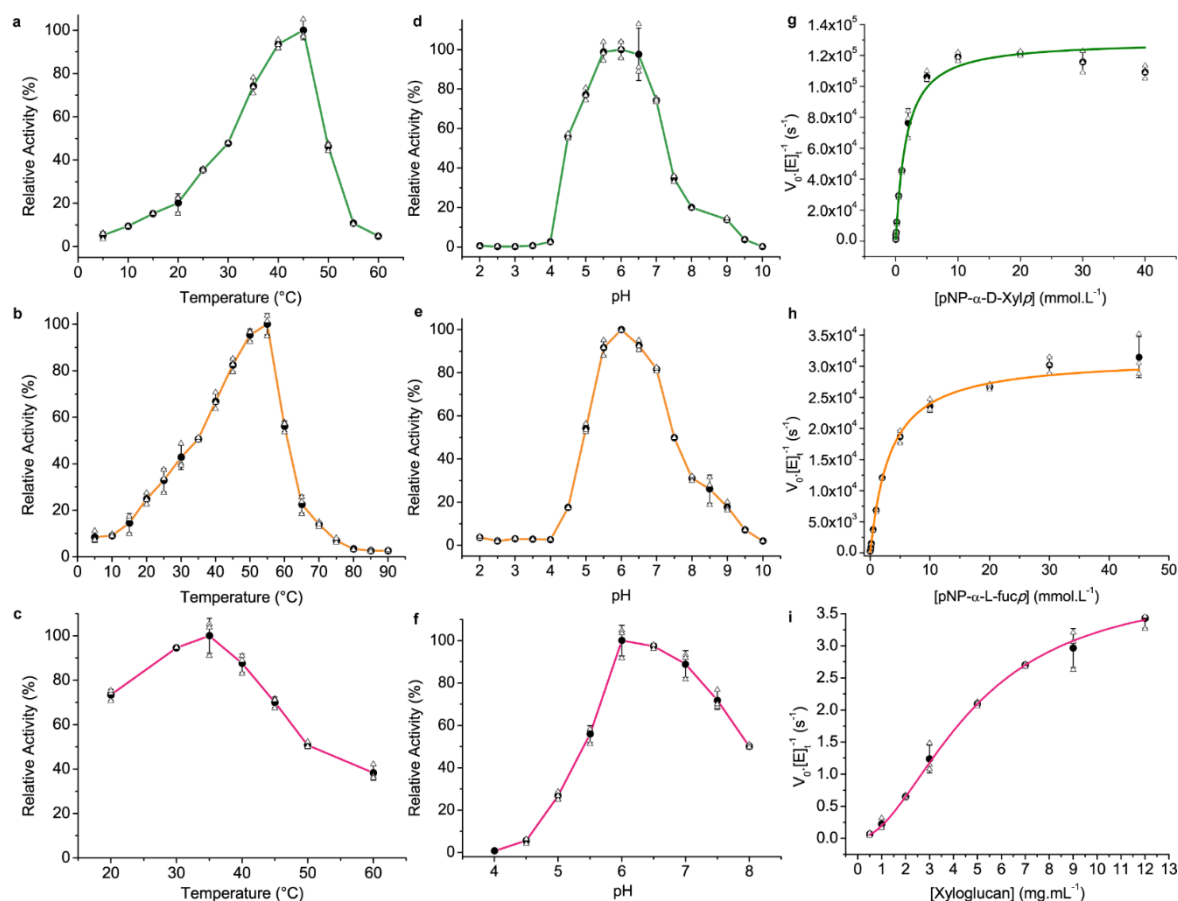
Supplementary Figure 10. CE20 xyloglucan acetylesterase *XacXaeA* structural analysis. (a) Active site residues represented in sticks. In brown, the C-atoms of the catalytic triad Asp-His-Ser, where the serine is covalently bound to PMSF (black C-atoms) used during the purification steps. C-atoms from the residues forming the oxyanion hole are shown in light green. Dashed lines represent the hydrogen bonds. (b) The electrostatic surface potential (generated using APBS⁷) of *XacXaeA* structure, colored from blue (positively charged) to red (negatively charged). The black circles delineate the active site. (c) Conservation score derived from ConSurf^{8,9}, projected over the structure surface representation, with highly conserved amino acid residues in dark purple, semi-conserved residues in white, and variable residues in dark green. (d to h) Representative crystallographic structures from other CAZy carbohydrate esterase families with SGNH-hydrolase fold (cartoon representation) and their respective electrostatic surface potential colored as in (b). CE6 (PDB ID 2APJ¹⁰) (d), CE2 (PDB ID 3U37¹¹) (e), CE17 (PDB ID 6HFZ¹²) (f), CE3 (PDB ID 2VPT¹³) (g) and CE12 (PDB ID 1DEO¹⁴) (h). (See also Fig. 3).



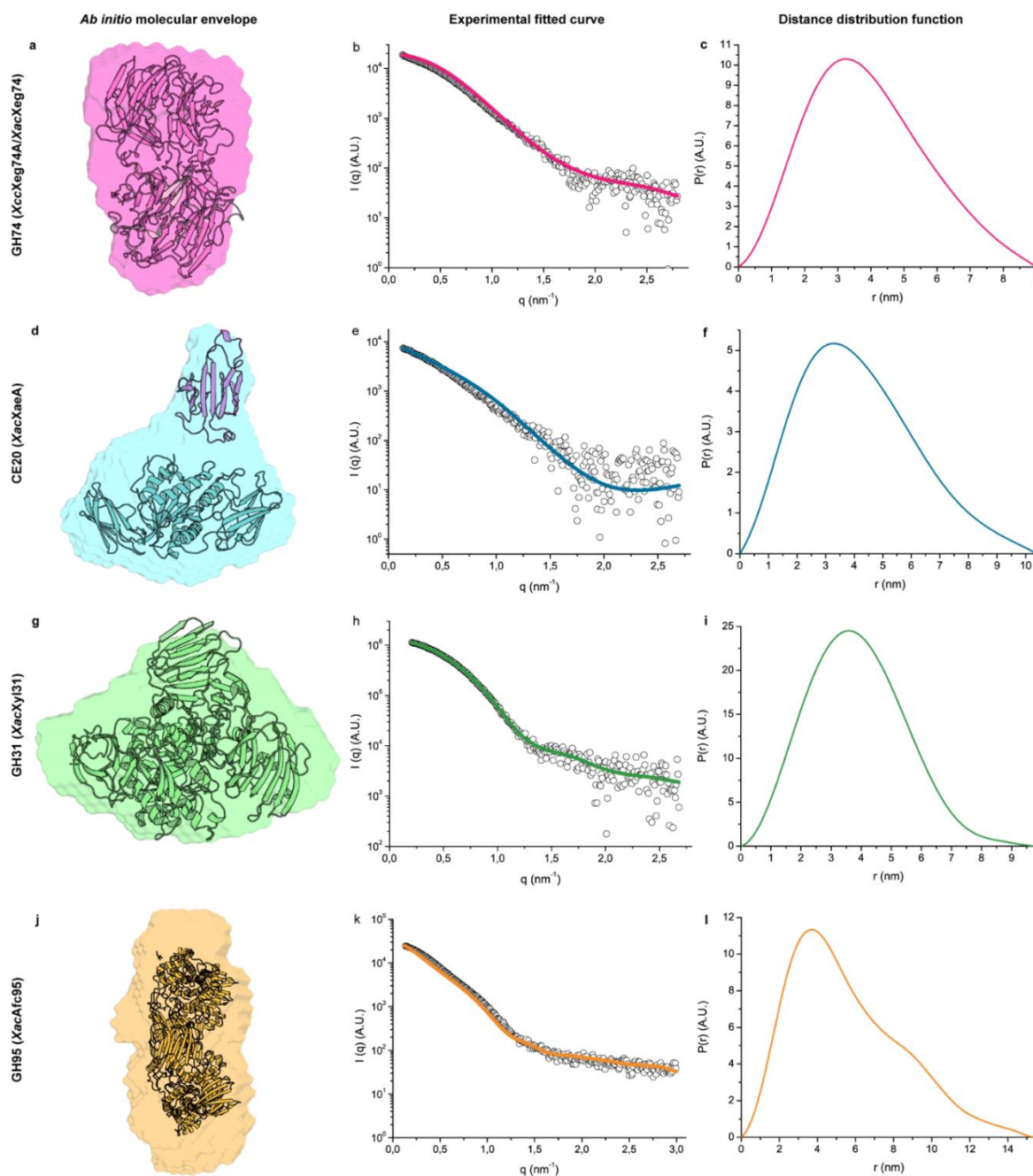
Supplementary Figure 11. Alignment of CE20 xyloglucan acetyltransferase sequences from representative *Xanthomonas* species. Sequence numbering based on *XacXaeA*. Secondary structure elements are represented and labeled according to the crystallographic structure of *XacXaeA*. Arrows = β -sheets, helices = α -helices and T = β -turn. Each box color represents a specific domain (See Fig. 3): Light green - the β -sandwich domains at the N- and C-termini, light blue - the catalytic core SGNH-hydrolase domain, and light pink - the X448 domain. Green and red circles indicate the residues that form the oxyanion hole and the catalytic triad, respectively. The four boxes delineate the characteristic SGNH hydrolase motifs known as Blocks I (purple), II (brown), III (dark green) and V (yellow). Residues shaded in black are fully conserved. The representative species are: *xac*, *Xanthomonas citri* pv. *citri* 306, *xcp*, *Xanthomonas campestris* pv. *raphani* 756C, *xcx*, *Xanthomonas cucurbitae* ATCC 23378, *xga*, *Xanthomonas gardneri* ICMP 7383, *xva*, *Xanthomonas vasicola* pv. *vasculorum* SAM119, *xph*, *Xanthomonas phaseoli* pv. *phaseoli* CFBP6546R, *xax*, *Xanthomonas axonopodis* pv. *citrumelo* F1, *xcv*, *Xanthomonas campestris* pv. *vesicatoria* 85-10, *xpe*, *Xanthomonas perforans* LH3, *xfu*, *Xanthomonas citri* pv. *fuscans* 4834-R. Alignment was generated with the Clustal Ω^1 server and the figure was produced using the ESPrnt 3.0 server².



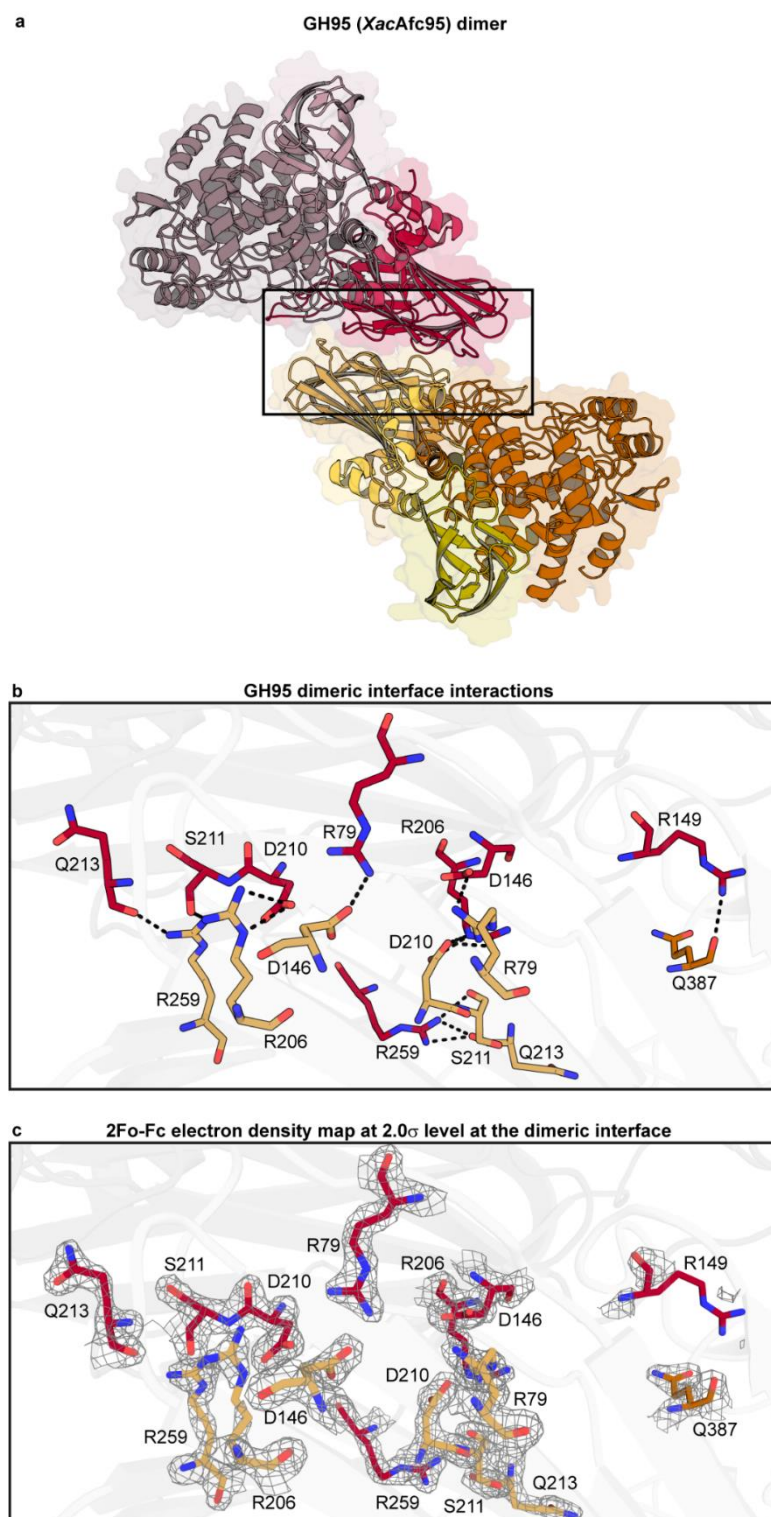
Supplementary Figure 11. Alignment of CE20 xyloglucan acetylerase sequences from representative *Xanthomonas* species (Continued). Sequence numbering based on *XacXaeA*. Secondary structure elements are represented and labeled according to the crystallographic structure of *XacXaeA*. Arrows = β -sheets, helices = α -helices and T = β -turn. Each box color represents a specific domain (See Fig. 3): Light green - the β -sandwich domains at the N- and C-termini, light blue - the catalytic core SGNH-hydrolase domain, and light pink - the X448 domain. Green and red circles indicate the residues that form the oxyanion hole and the catalytic triad, respectively. The four boxes delineate the characteristic SGNH hydrolase motifs known as Blocks I (purple), II (brown), III (dark green) and V (yellow). Residues shaded in black are fully conserved. The representative species are: *xac*, *Xanthomonas citri* pv. *citri* 306, *xcp*, *Xanthomonas campestris* pv. *raphani* 756C, *xcz*, *Xanthomonas cucurbitae* ATCC 23378, *xga*, *Xanthomonas gardneri* ICMP 7383, *xva*, *Xanthomonas vasicola* pv. *vasculorum* SAM119, *xph*, *Xanthomonas phaseoli* pv. *phaseoli* CFBP6546R, *xax*, *Xanthomonas axonopodis* pv. *citrumelo* F1, *xcv*, *Xanthomonas campestris* pv. *vesicatoria* 85-10, *xpe*, *Xanthomonas perforans* LH3, *xfu*, *Xanthomonas citri* pv. *fuscans* 4834-R. Alignment was generated with the Clustal Ω^1 server and the figure was produced using the ESPript 3.0 server².



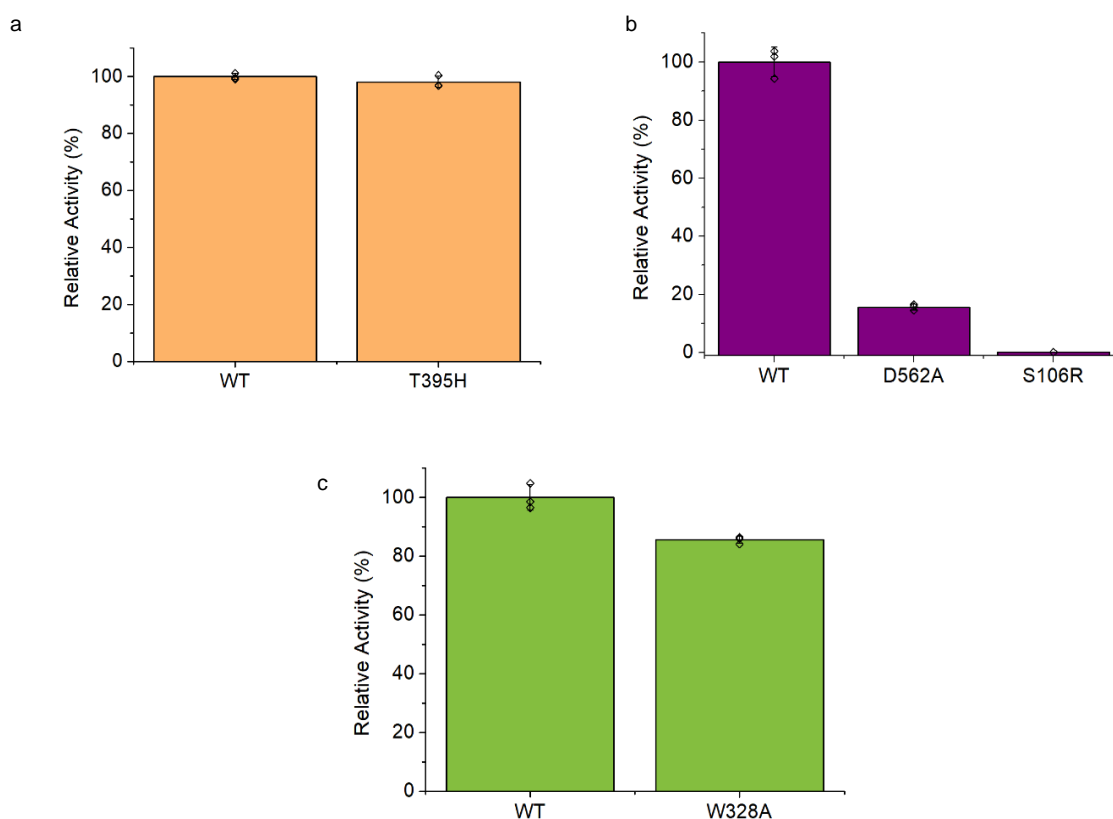
Supplementary Figure 12. Biochemical and kinetic characterization of XyGUL enzymes. *Left*, temperature dependence curves of XacXyl31 (a), XacAfc95 (b) and XacEgl9 (c). *Middle*, pH dependence curves of XacXyl31 (d), XacAfc95 (e) and XacEgl9 (f). *Right*, substrate saturation curves of XacXyl31 using *para*-nitrophenyl- α -D-xylopyranoside (g), XacAfc95 using *para*-nitrophenyl- α -L-fucopyranoside (h) and XacEgl9 using tamarind xyloglucan (i). Curves are colored according to the enzyme: XacXyl31 (green), XacAfc95 (orange), XacEgl9 (pink). Data are shown as mean \pm SD (three independent experiments, $n=3$). (V_0 =initial velocity; $[E]_t$ =enzyme concentration). Mean values are shown as filled circles, while measured data are shown as empty circles. Standard deviations are shown as black lines. Source data are provided as a source data file.



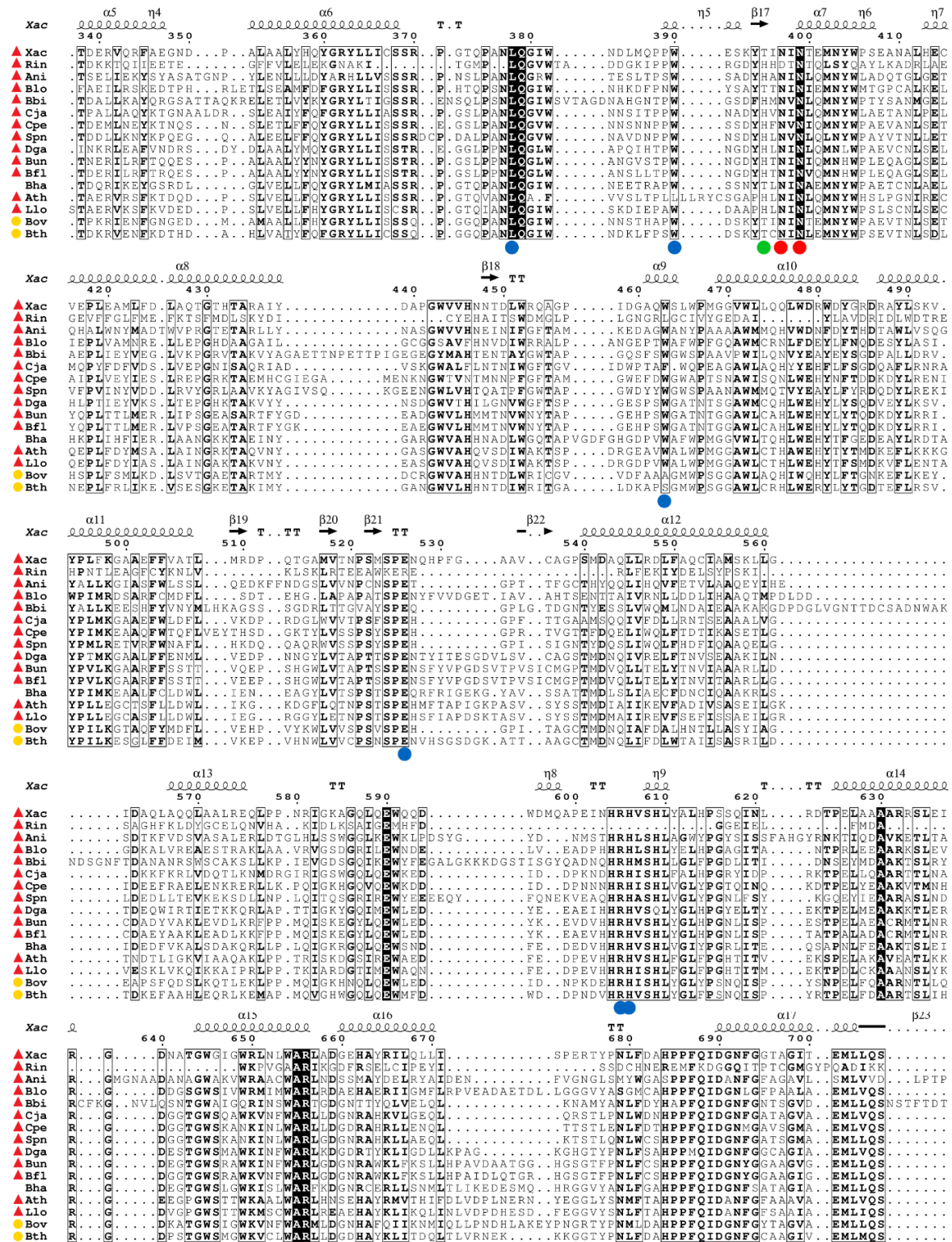
Supplementary Figure 13. SAXS analysis of XyGUL enzymes. *Left*, crystal structures (represented as cartoons) fitted into the SAXS envelopes for *XacXeg74* monomer (in pink), (with crystallographic structure of homologue *XccXeg74*, SeqId 84%) (a), *XacXaeA* monomer with the core module in a blue cartoon and the homology-modeled X448 domain in light pink (d), *XacXyl31* monomer in green) (g) and *XacAfc95* dimer in orange) (j). *Middle*, Experimental (empty circles) and calculated (lines) scattering curves of each enzyme in solution: *XacXeg74* (b), *XacXaeA* (e), *XacXyl31* (h) and *XacAfc95* (k). *Right*, Distance distribution curves used for envelope generation and theoretical scattering calculation based on each experimental curve: *XacXeg74* (c), *XacXaeA* (f), *XacXyl31* (i) and *XacAfc95* (l).



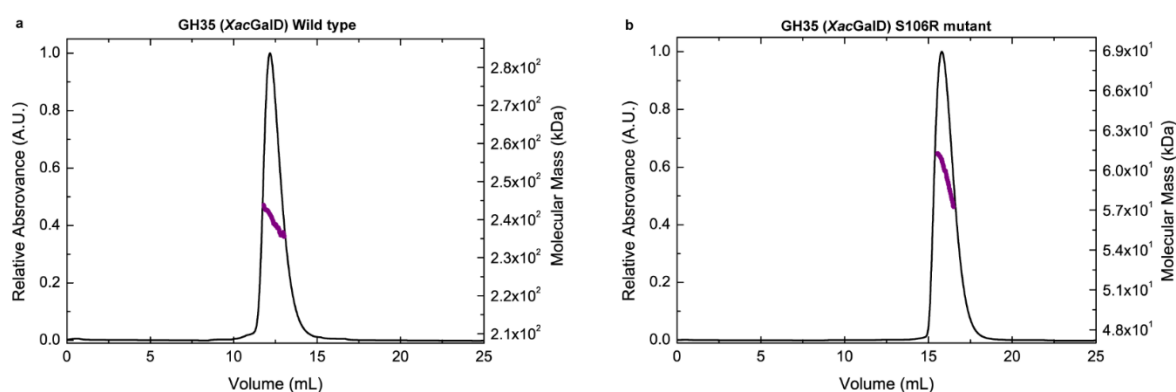
Supplementary Figure 14. GH95 *XacAfc95* oligomeric interface. (a) Cartoon representation of the *XacAfc95* dimer with C-atoms of each protomer colored in shades of pink and yellow, respectively, to highlight the domains. The rectangle delimits the oligomeric interface. (b) Residues forming the dimeric interface are labeled and represented as sticks with C carbons colored as in (a). Dashed lines represent hydrogen bonds. (c) Dimeric interface as in (b), with 2F_o-F_c electron density map contoured at 2σ level.



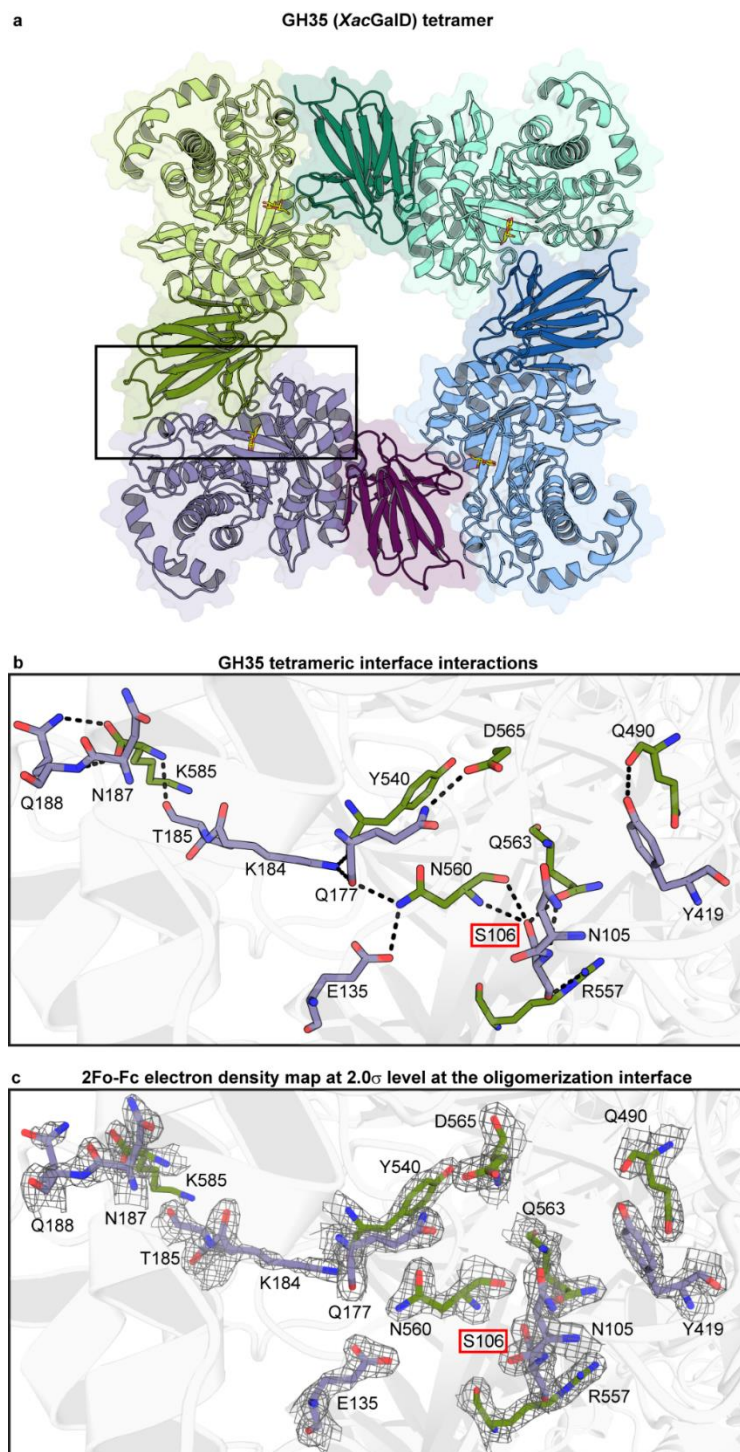
Supplementary Figure 15. Effect of point mutations on the activity of XyGUL enzymes. (a) Wild-type *XacAfc95* (WT) and mutant T395H tested with *para*-nitrophenyl- α -L-fucopyranoside. (b) Wild-type *XacGalD* and mutants D562A and S106R tested with *para*-nitrophenyl- β -D-galactopyranoside. (c) Wild-type *XacXyl31* and mutant W328A tested with *para*-nitrophenyl- α -D-xylopyranoside. Data are shown as mean \pm SD from three independent experiments (n=3). Measured data are shown as empty circles. Standard deviations are shown as black lines. Source data are provided as a data file.



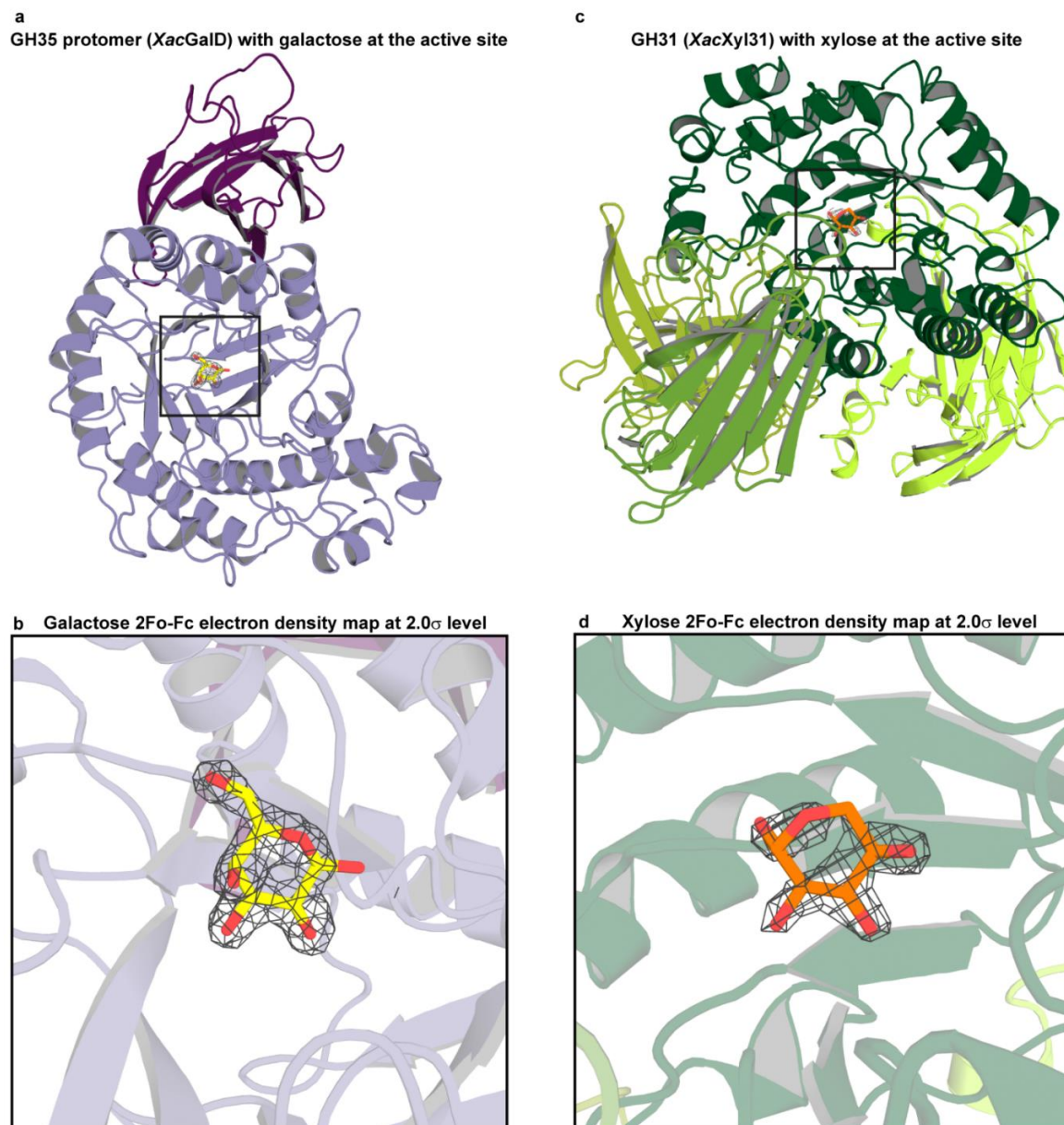
20088, *Bbi*, *Bifidobacterium bifidum* JCM 1254, *Cja*, *Cellvibrio japonicus* Ueda107, *Cpe*, *Clostridium perfringens* ATCC 13124, *Spn*, *Streptococcus pneumoniae* TIGR4, *Dga*, *Dysgonomonas gadei* ATCC BAA-286, *Bun*, *Bacteroides uniformis* ATCC 8492, *Bfl*, *Bacteroides fluxus* YIT 12057, *Bha*, *Bacillus halodurans* C-125, *Ath*, *Arabidopsis thaliana*, *Llo*, *Lilium longiflorum*, *Bov*, *Bacteroides ovatus* ATCC 8483, *Bth*, *Bacteroides thetaiotaomicron* VPI-5482. Alignment performed with Clustal Ω^1 server and image produced with ESPript 3.0 server².



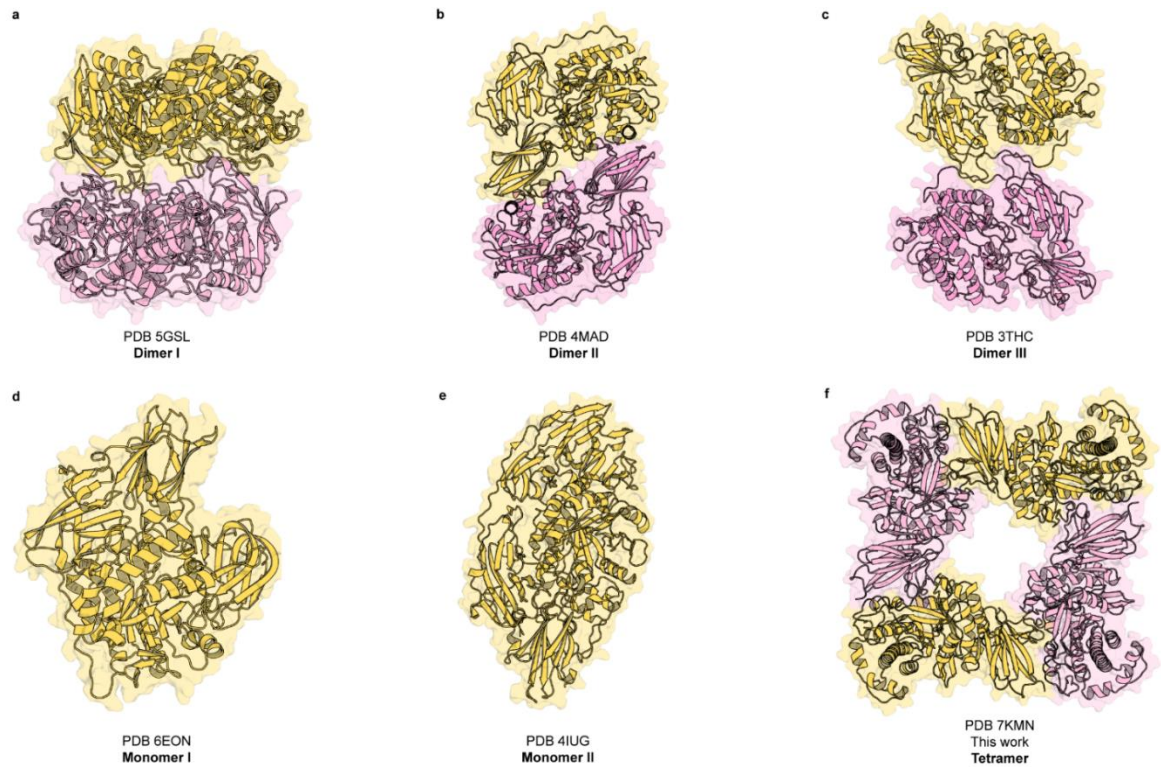
Supplementary Figure 17. Mutation S106R disrupts the GH35 *XacGalD* tetramer. Size-exclusion chromatography with multiple-angle light scattering (SEC-MALS) analysis of (a) wild-type *XacGalD* (tetramer) and (b) S106R mutant (monomer). Black curves represent the relative absorbance at 280 nm of the proteins eluted from a Superdex 200 HR 10/300 GL analytical size-exclusion column. Purple lines represent the molecular mass of the eluted proteins calculated from the multiple-angle light scattering measurements.



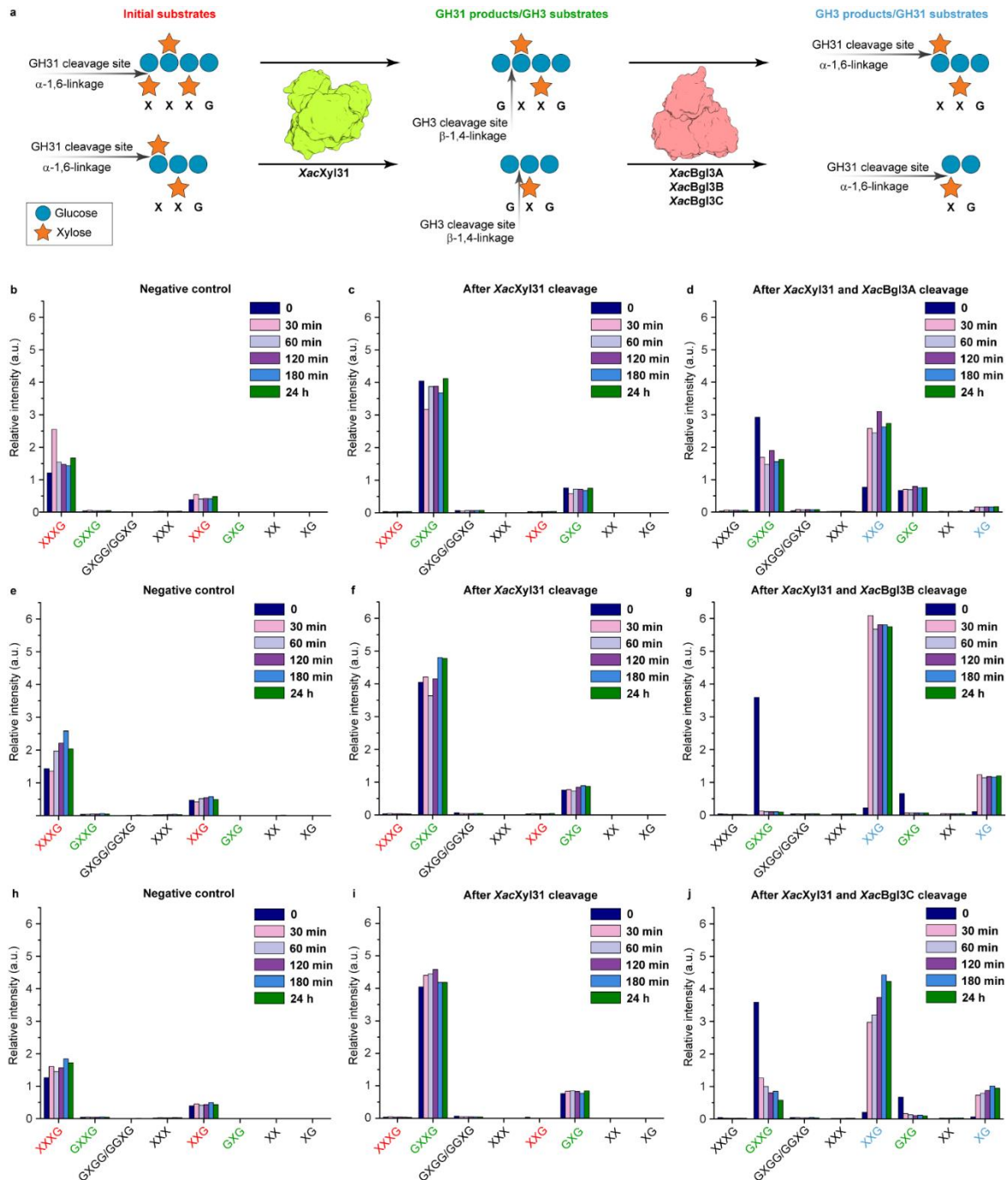
Supplementary Figure 18. GH35 *XacGalD* tetramer. (a) Crystal structure of *XacGalD* tetramer (cartoon) with each protomer represented in shades of a distinct color to highlight the catalytic domain (light shades) and the β -sandwich domain (dark shades). Sticks represent D-galactose (yellow C-atoms). The rectangle delimits the oligomeric interface, repeated between every two protomers. (b) Amplified view of the oligomeric interface with residues represented as sticks and C-atoms colored as in (a). Dashed lines represent hydrogen bonds. The residue S106, which disrupts the interface when mutated to arginine, is marked with a red box. (c) Oligomeric interface as in (b), with $2F_o - F_c$ electron density map contoured at 2σ level.



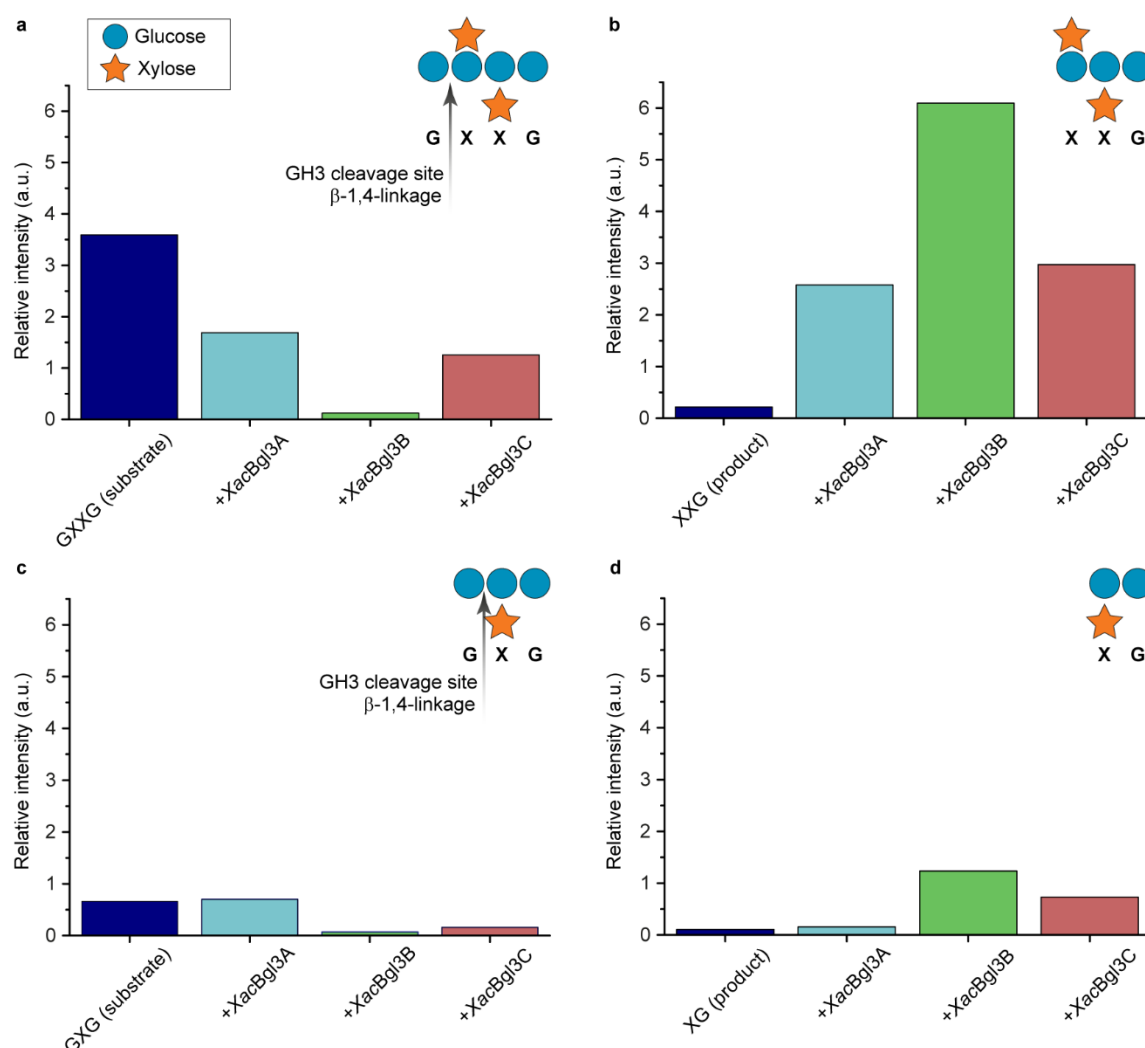
Supplementary Figure 19. GH35 *XacGalD* and GH31 *XacXyl31* crystallographic structures in complex with products. *XacGalD* (a) and *XacXyl31* (c) protomers are represented in cartoon, with the bound galactose (yellow C-atoms) or xylose (orange C-atoms) shown as sticks, respectively. Domains are colored in shades of purple in *XacGalD* or green in *XacXyl31*. The 2Fo-Fc electron density maps contoured at 2 σ level of galactose (b) and xylose (d) in the active site of *XacGalD* and *XacXyl31*, respectively.



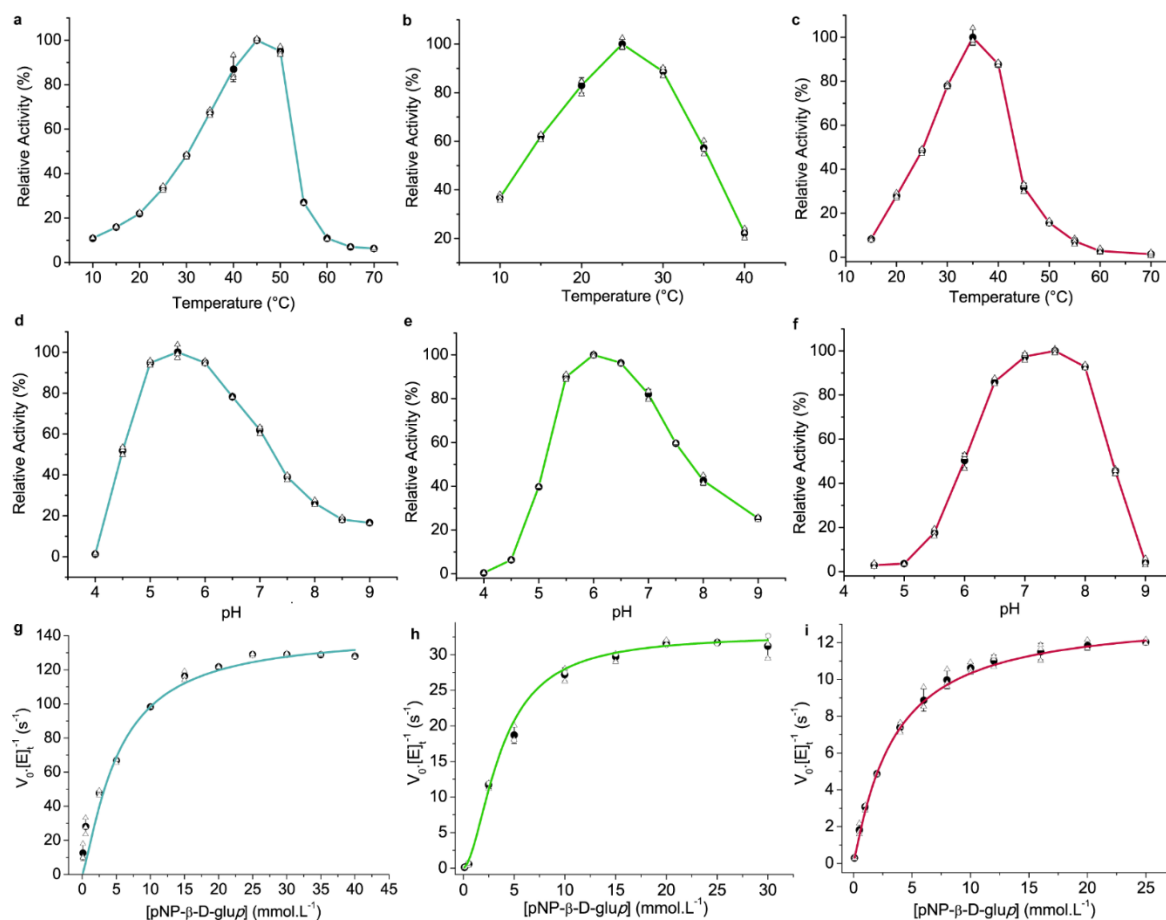
Supplementary Figure 20. The diversity of oligomeric arrangements in the GH35 family. Each protomer (or monomer) is represented in yellow or pink cartoon and transparent surface. Structural analysis is also provided in Supplementary Table 11. (a) Archaea - *Pyrococcus horikoshii* OT3 (PDB 5GSL) dimer interface type I, (b) Firmicutes - *Bacillus circulans* ATCC 31382 (PDB 4MAD¹⁵) dimer interface type II, (c) Chordate - *Homo sapiens* (PDB 3THC¹⁶) dimer interface type III, (d) Bacteroidetes - *Bacteroides thetaiotaomicron* VPI-5482 (PDB 6EON¹⁷) monomer type I (with 3 accessory domains), (e) Fungi - *Aspergillus oryzae* RIB40 (PDB 4IUG¹⁸) monomer type II (with 4 accessory domains), and (f) Proteobacteria - *Xanthomonas citri* pv. *citri* 306 tetramer (this work, Fig. 4).



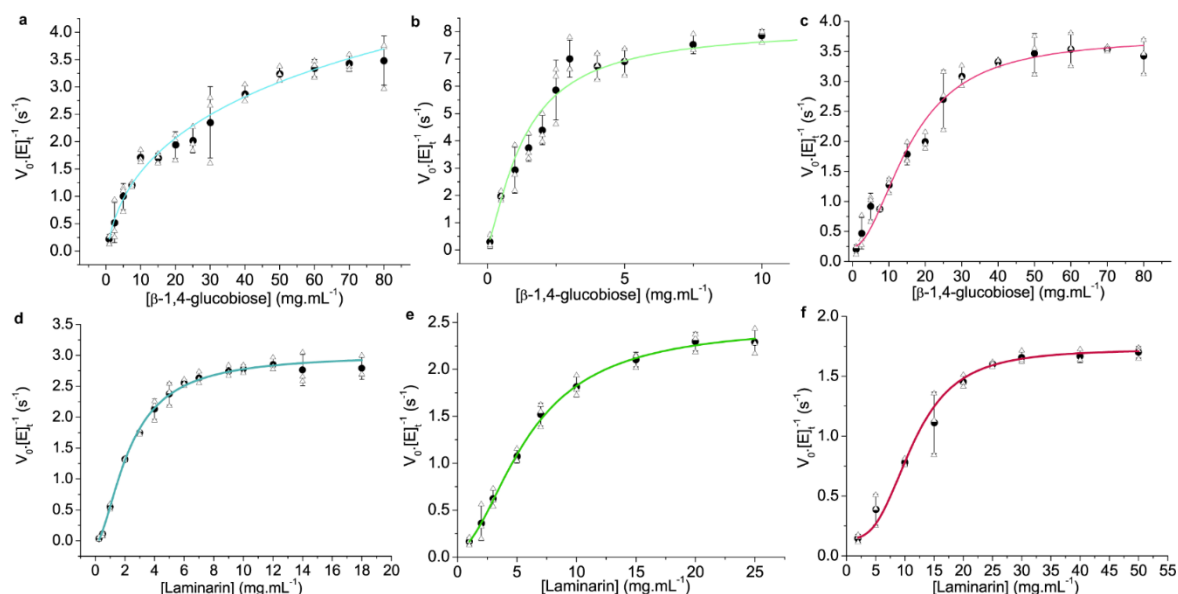
Supplementary Figure 21. GH3 β -glucosidases from *X. citri* are active on xyloglucan oligosaccharides pre-treated with *XacXyl31*. (a) Schematic representation of the XyGOs used as substrates (*left*) for *XacXyl31* (green surface), which removes the xylosyl moiety at the non-reducing end, generating products (*middle*) that are substrates for GH3 β -glucosidases (pink surface), which remove the glycosyl moiety at the non-reducing end from the backbone, releasing shorter substrates for the further action of *XacXyl31* (*right*). (b-j) Initial substrates (red labels), *XacXyl31* products (green labels), and GH3 products (cyan labels) detected by mass spectrometry at several incubation times without enzyme (negative control, b, e and h), *with XacXyl31* (c, f and i), and after thermal inactivation of *XacXyl31* and subsequent incubation with each one of the three GH3 β -glucosidases from *X. citri* (d, g and j). The relative intensity of the peaks was normalized using an internal standard. Carbohydrates are represented by geometric shapes (glucose: blue circles and xylose: orange stars).



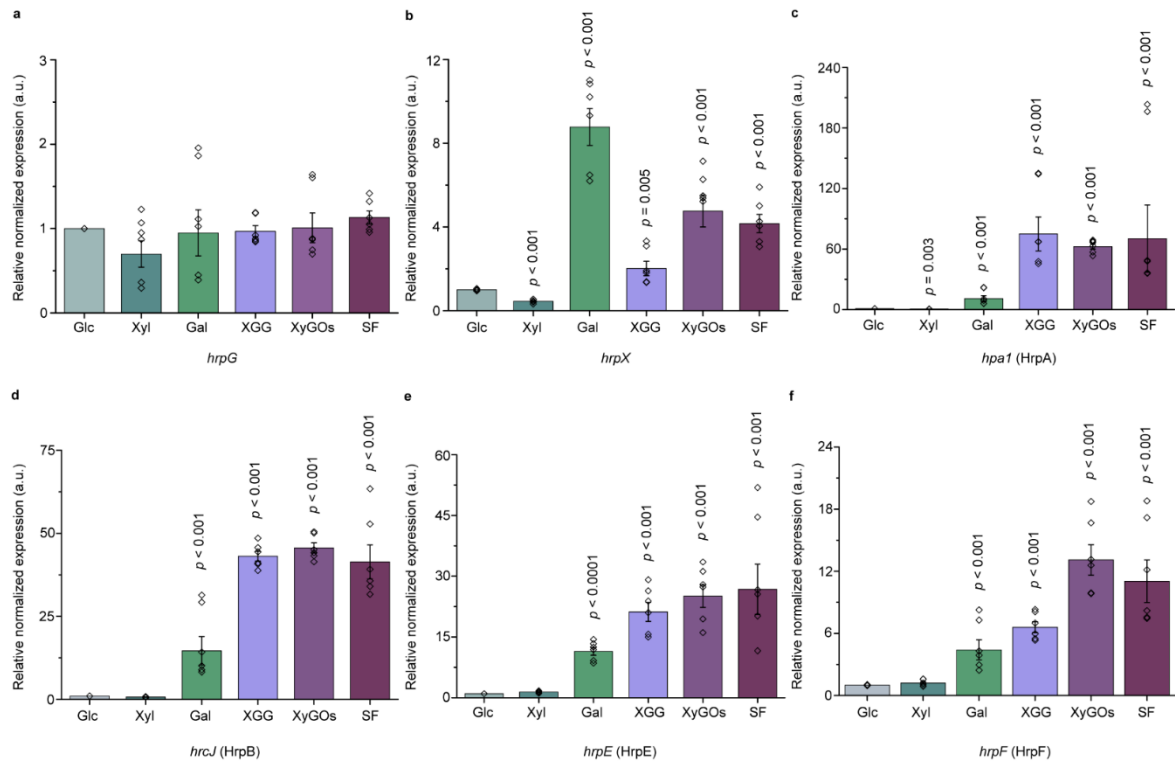
Supplementary Figure 22. A simplified view of the results shown in the previous Supplementary Fig. 21, highlighting substrate depletion (a and c) and product generation (b and d) after 30 min incubation with 100 μ g of GH3 β -glucosidases from *X. citri*. The negative control (reaction without enzyme) is shown in blue and the other reactions are colored according to the respective enzyme indicated in the X axis. The relative intensity of the peaks was normalized using an internal standard. Carbohydrates are represented by geometric shapes (glucose: blue circles and xylose: orange stars).



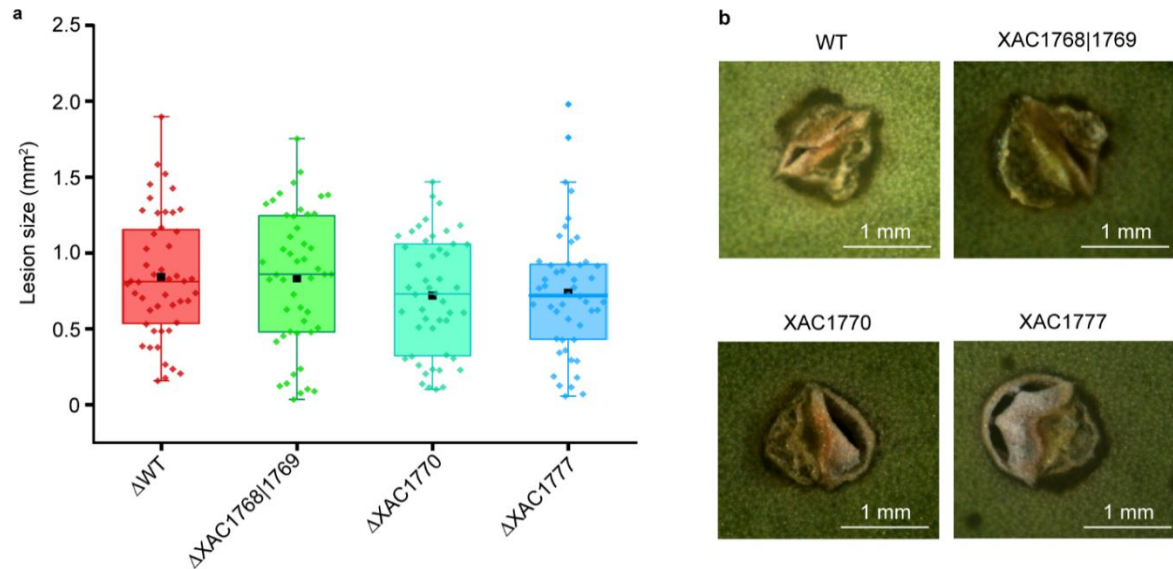
Supplementary Figure 23. Biochemical characterization of the GH3 β -glucosidases *XacBgl3A*, *XacBgl3B* and *XacBgl3C*. Temperature dependence curves of *XacBgl3A* (a), *XacBgl3B* (b) and *XacBgl3C* (c). pH dependence curves of *XacBgl3A* (d), *XacBgl3B* (e) and *XacBgl3C* (f). Substrate saturation curves using *para*-nitrophenyl- β -D-glucopyranoside for *XacBgl3A* (g), *XacBgl3B* (h) and *XacBgl3C* (i). The curves are colored according to the enzyme: *XacBgl3A* (light blue), *XacBgl3B* (green) and *XacBgl3C* (red). Data are shown as mean \pm SD from three independent experiments ($n=3$). (V_0 =initial velocity and $[E]_i$ =enzyme concentration). Mean values are shown as filled circles, while measured data are shown as empty circles. Standard deviations are shown as black lines. Source data are provided as a data file.



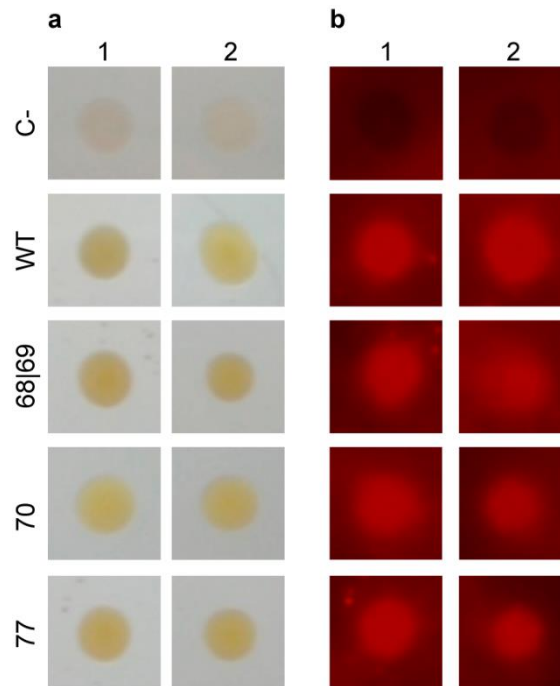
Supplementary Figure 24. Biochemical characterization of the GH3 β -glucosidases *XacBgl3A*, *XacBgl3B* and *XacBgl3C*. Substrate saturation curves using β -1,4-glucobiose for *XacBgl3A* (a), *XacBgl3B* (b) and *XacBgl3C* (c). Substrate saturation curves using polymeric laminarin for *XacBgl3A* (d), *XacBgl3B* (e) and *XacBgl3C* (f). Note that these enzymes cleave both β -1,4 and β -1,3 linkages between glucosyl residues from cellooligosaccharides (β -1,4) or laminarin (β -1,3), indicating they might play a role in the final steps not only of XyG or cellulose depolymerization but also of other glucans such as callose (β -1,3 glucan backbone). The curves are colored according to the enzyme: *XacBgl3A* (light blue), *XacBgl3B* (green) and *XacBgl3C* (pink). Data are shown as mean \pm SD from three independent experiments ($n=3$). (V_0 =initial velocity and $[E]_t$ =enzyme concentration). Mean values are shown as filled circles, while measured data are shown as empty circles. Standard deviations are shown as black lines. Source data are provided as a data file.



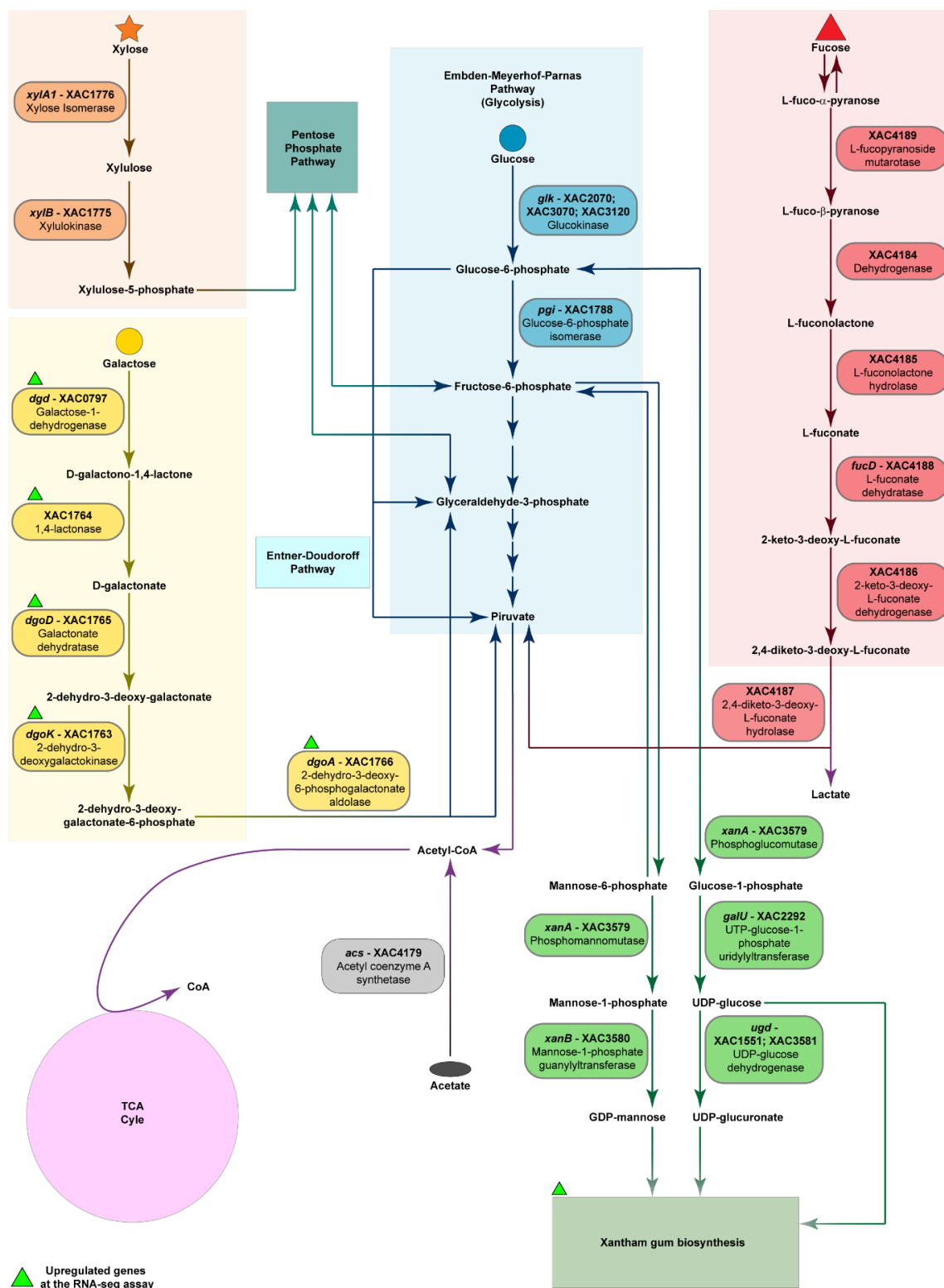
Supplementary Figure 25. Expression profile of *hrpG*, *hrpX* and representative genes of the *hrp* cluster of *X. citri* under different culture conditions. RT-qPCR data analysis showing the normalized relative expression of *hrpG*, *hrpX*, *hpa1*, *hrcJ*, *hrpE* and *hrpF* genes in a minimal medium XVM2¹⁹, termed here as SF (saccharose plus fructose), or modified XVM2 in which sucrose and fructose were replaced by different carbohydrates: Glc, glucose; Xyl, xylose; Gal, galactose; XGG, sugar mix of xylose, galactose and glucose; and XyGOs, xyloglucan oligosaccharides. Bars are colored according to the respective condition indicated in the X axis. Values were normalized using two experimentally validated reference genes (XAC4218 and XAC2177). Data are shown as mean values (\pm SEM) from three independent biological experiments with technical duplicates. *p*-values indicate significant differences compared to the Glc condition as determined by two-tailed *t*-test. The operons corresponding to each gene are indicated in parentheses. Measured data are shown as empty circles. SEM values are shown as black lines. Source data are provided as a source data file.



Supplementary Figure 26. The knockout of selected XyGUL genes is insufficient to abolish virulence. (a) Box plot showing the area of canker lesions resulting from pinprick inoculations with *X. citri* in *Citrus sinensis* leaves imaged at 12 days post-infection. Each box plot presents the median (horizontal line inside the box), the mean value (black square), the first (Q_1) and third (Q_3) quartiles (lower and upper edges of the box), and whiskers (vertical lines) extending to the most extreme values inside lower [$Q_1 - 1.5(Q_3 - Q_1)$] and upper [$Q_3 + 1.5(Q_3 - Q_1)$] fences. Dots represent measured data and those outside the whiskers are outliers. (b) A representative image of the lesion promoted by each strain. WT = wild-type. The assays were performed with three independent biological samples, each composed of 16 technical replicates. Tukey test indicates no significant difference between knockout and WT strains at 0.05 level.

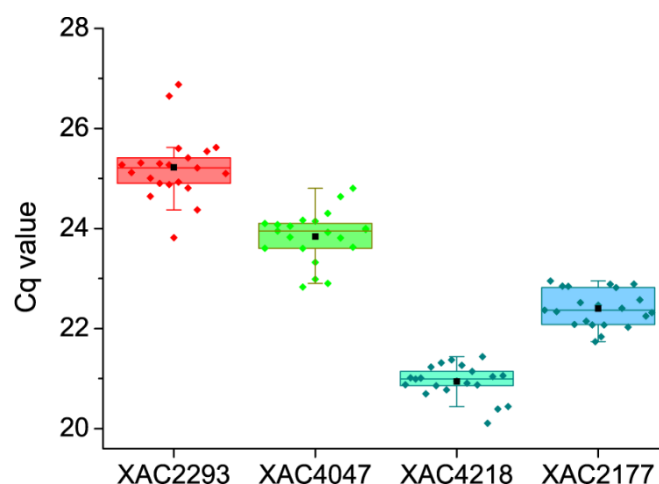


Supplementary Figure 27. The deletion of *XacXeg74* does not abolish the xyloglucanase activity of *X. citri*. (a) Colonies of *X. citri* strains and *E. coli* (C-) growth during 30 h in NYG medium containing 0.5% tamarind seed xyloglucan. 68|69= Δ XAC1768-69 (TonB-dependent transporters); 77= Δ XAC1777 (MFS transporter) and 70= Δ XAC1770 (GH74 xyloglucanase). (b) The same plate showed in panel (a) stained with congo red after removal of bacterial colonies to evidence the clearer halo corresponding to xyloglucan degradation. Note that the xyloglucan degradation halo is absent in the negative control (C-) and is not affected by the deletion of XyGUL genes, including XAC1770, which encodes a GH74 xyloglucanase. The plate is representative of three independent assays.

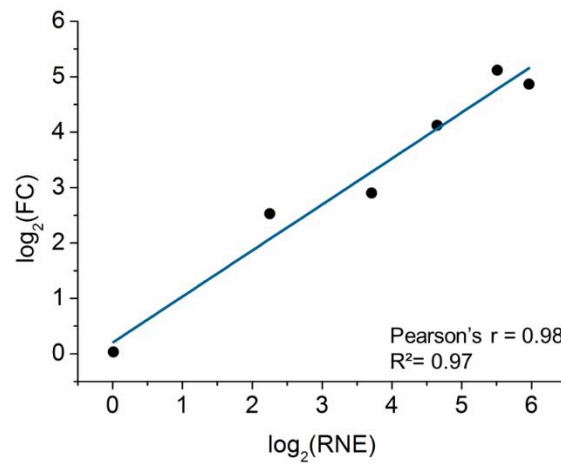


Supplementary Figure 28. Schematic representation of *X. citri* metabolism for XyG-released products. In the *X. citri* cytoplasm, monosaccharides released from XyG can be metabolized by specific pathways (boxes colored in light shades of orange, red, blue and yellow) that converge to the pentose phosphate pathway or to the tricarboxylic acid (TCA) cycle. Intermediate metabolites from glycolysis can also be converted into precursors for xanthan gum biosynthesis. The acetate released by the xyloglucan acetyltransferase *XacXaeA* is possibly imported to the cytoplasm through the cation/acetate

symporter *ppA* (XAC4176) and then converted in acetyl-CoA to enter the TCA cycle. The reconstruction of *X. citri* main reactions of the central metabolism was based on previous works with *X. campestris* pv. *campestris* B100^{20–23} and *X. campestris* pv. *campestris* ATCC 33913²⁴. Arrows represent enzyme-catalyzed metabolic interconversions. Rounded corner rectangles highlight the genes and respective enzymes associated with peripheral reactions. Co-factors were omitted to improve clarity. Carbohydrates and acetate are represented by geometric shapes (glucose: blue circle, xylose: orange star, fucose: red triangle, galactose: yellow circle and acetate: gray ellipse). Green triangles indicate genes upregulated in the XyGOs condition, as identified at the RNA-seq assay presented in this work.



Supplementary Figure 29. Variation in the expression of candidate reference genes. Each box plot of Cq data (dots) presents the median (horizontal line inside the box), the mean value (black square), the first (Q_1) and third (Q_3) quartiles (lower and upper edges of the box), and whiskers (vertical lines) extending to the most extreme values inside lower [$Q_1 - 1.5(Q_3 - Q_1)$] and upper [$Q_3 + 1.5(Q_3 - Q_1)$] fences. Dots outside the whiskers are considered outliers. The variation in the expression levels was evaluated in seven different culture conditions for each gene. Statistics are derived from three independent experiments with technical duplicates. Cq: quantification cycle.



Supplementary Figure 30. Pearson's correlation between RNA-seq and RT-qPCR data. $\log_2(\text{FC})$ values from RNA-seq assays and $\log_2(\text{RNE})$ values from RT-qPCR assays were analyzed for 6 genes (black dots) using Pearson's correlation. r = Pearson correlation coefficient. R^2 = coefficient of determination (p -value < 0.001 , two-tailed). Blue line = linear regression fit. Gene expression data were obtained for *X. citri* growth in minimal medium XVM2m containing XyGOs in comparison to XVM2m glucose medium. FC: fold change and RNE: relative normalized expression.

Supplementary Table 1. Sequence-based prediction of subcellular localization of proteins from the *Xanthomonas* XyGUL and accessory enzymes for xyloglucan depolymerization. Of note, the position of the start codon originally annotated for these ORFs was revised based on the reference sequences whose accession codes are shown in NR ID. ^aSPII signal predicts for lipoproteins, *i.e* membrane-tethered proteins via N-terminal lipidation. ^bPosition +2 different from D (aspartic acid) after a SPII cleavage site indicates outer-membrane anchoring²⁵. TMH = transmembrane helix. CYT = cytoplasm, EXT = extracellular, IM = inner membrane, PER = periplasm, OM = outer membrane.

Enzymes encoded by the XyGUL								
Gene	Protein	NR ID	Description	SignalP 5.0 ²⁶	TMHMM ²⁷	CELLO v.2.5 ²⁸	Consensus	Criterion
XAC1770	<i>XacXeg74</i>	WP_011051128.1	Endo- β -1,4-Xyloglucanase	Sec/SPI (1-38)	1 TMH (13-32)	EXT	EXT	SPI site accessible
XAC1771	<i>XacXaeA</i>	WP_011051129.1	Xyloglucan Acetylerase	Sec/SPI (1-23)	1 TMH (9-31)	PER	PER	SPI site accessible (region 24-31 is seen in the crystal structure and is not helical)
XAC1772	<i>XacGalD</i>	WP_033485297.1	β -Galactosidase	Sec/SPII ^a (1-34) +2=Q	-	PER	OM	N-terminal lipid anchor / +2 \neq D ^b
XAC1773	<i>XacXyl31</i>	WP_011051131.1	α -Xylosidase	Sec/SPI (1-36)	-	PER	PER	SPI site accessible/PER
XAC1774	<i>XacAfc95</i>	WP_011051132.1	α -L-1,2-Fucosidase	Tat/SPI (1-27)	-	PER	PER	SPI site accessible/PER
Putative accessory enzymes encoded outside the XyGUL								
Gene	Protein	NR ID	Description	SignalP 5.0 ²⁶	TMHMM ²⁷	CELLO v.2.5 ²⁸	Consensus	Criterion
XAC0029	<i>XacEgl5B</i>	WP_011050033.1	Endo- β -1,4-glucanase	Sec/SPI (29-30)	-	OM	EXT ²⁹	SPI site accessible and literature
XAC1448	<i>XacBgl3A</i>	WP_040107594.1	β -Glucosidase	Sec/SPI (1-16)	-	PER	PER	SPI site accessible/PER
XAC1793	<i>XacBgl3B</i>	WP_011051143.1	β -glucosidase	Sec/SPII ^a (1-30) +2=Q	-	PER	OM	N-terminal lipid anchor pos+2 \neq D ^b
XAC2522	<i>XacEgl9</i>	WP_011051647.1	Endo- β -1,4-glucanase	Sec/SPII [¥] (1-19) +2=A	-	PER	OM	N-terminal lipid anchor pos+2 \neq D ^b
XAC3869	<i>XacBgl3C</i>	WP_011052578.1	β -Glucosidase	-	-	CYT	CYT	Lack of signal peptide
XAC4183	<i>XacAbf43A</i>	WP_011052791.1	α -L-arabinofuranosidase	TAT/SPI (1-30)	-	PER	PER	SPI site accessible/PER

Supplementary Table 2. BLAST search using *XacXaeA* sequence as query against the Protein Data Bank database. Note the low sequence coverage along with the low sequence identity.

Description	PDB code	Sequence identity (%)	E-value	Sequence Coverage (%)
GH2 β -glucuronidase from <i>Bacteroides uniformis</i>	6D8G ³⁰	32.63	2e-05	14
GH2 β -Galactosidase from <i>Thermotoga maritima</i>	6SD0 ³¹	27.12	0.017	23
GH2 β -Galactosidase from <i>Bacillus circulans</i>	4YPJ ³²	28.57	0.072	15
GH2 β -Galactosidase from <i>Bacteroides vulgatus</i>	3GM8	24.03	0.23	17
GH2 β -Glucuronidase from <i>Ruminococcus gnavus</i>	6JZ1	30.14	0.27	11
CE6 member from <i>Arabidopsis thaliana</i>	2APJ ¹⁰	26.05	0.35	17

Supplementary Table 3. List of *Xanthomonas* species and information about the disease and tissue/host specificity.

Specie	Host	Disease	Local of infection	Reference
<i>X. campestris</i> pv. <i>campestris</i> 3811	Brassicaceae; <i>Arabidopsis thaliana</i>	Black rot	Vascular	33,34
<i>X. campestris</i> pv. <i>raphani</i> 756C	Brassicaceae; <i>Arabidopsis thaliana</i>	Bacterial spot	Non-vascular	35,36
<i>X. vesicatoria</i> ATCC 35937 LMG911	Tomato and pepper	Bacterial spot	Non-vascular	37–39
<i>X. vesicatoria</i> LM159	Tomato and pepper	Bacterial spot	Non-vascular	37–39
<i>X. cucurbitae</i> ATCC 23378	Pumpkin	Bacterial spot	Non-vascular	39–41
<i>X. campestris</i> pv. <i>badrii</i> NEB122	<i>Xanthium strumarium</i> , <i>Pisum sativum</i>	Unknown	Unknown	42
<i>X. hortorum</i> B07-007	Lettuce (<i>Lactuca sativa</i>)	Bacterial spot	Unknown	43
<i>X. arboricola</i> 17	Unknown	Unknown	Unknown	-
<i>X. arboricola</i> pv. <i>pruni</i> 15-088	<i>Prunus</i> species	Bacterial spot	Non-vascular	44,45
<i>X. arboricola</i> pv. <i>juglandis</i>	Walnut	Walnut blight	Non-vascular	45
<i>X. sp.</i> CPBF 424	Walnut	Walnut blight	Unknown	46
<i>X. fragariae</i>	Strawberry	Bacterial angular spot	Non-vascular	39,47,48
<i>X. axonopodis</i> pv. <i>commiphoreae</i> LMG26789	Guggal (<i>Commiphora wightii</i>)	Gumming of Guggal	Unknown	49
<i>X. axonopodis</i> pv. <i>citrumelo</i> F1	<i>Citrus</i> species	Bacterial spot	Non-vascular	50,51
<i>X. campestris</i> pv. <i>vesicatoria</i> str. 85-10	Pepper	Bacterial spot	Non-vascular	52
<i>X. euvesicatoria</i> LMG930	Tomato	Bacterial spot	Non-vascular	37–39
<i>X. perforans</i> 91-118	Tomato	Bacterial spot	Non-vascular	39,53
<i>X. phaseoli</i> pv. <i>phaseoli</i> CFBP412	Common bean (<i>Phaseolus vulgaris</i>)	Bacterial blight	Vascular	54
<i>X. citri</i> pv. <i>fuscans</i> 4834-R	Common bean (<i>Phaseolus vulgaris</i>)	Bacterial blight	Vascular	55
<i>X. citri</i> pv. <i>phaseoli</i> var. <i>fuscans</i> CFBP4885*	Common bean (<i>Phaseolus vulgaris</i>)	Bacterial blight	Vascular	55,56
<i>X. citri</i> pv. <i>vignicola</i> CFBP7113	Cowpea (<i>Vigna unguiculata</i>)	Bacterial blight	Non-vascular	57,58
<i>X. citri</i> pv. <i>aurantifolii</i> 1566	<i>Citrus</i> species	Citrus Canker	Non-vascular	39,59
<i>X. citri</i> pv. <i>citri</i> Aw12879	<i>Citrus</i> species	Citrus Canker	Non-vascular	39,60
<i>X. axonopodis</i> Xac29-1	<i>Citrus</i> species	Citrus Canker	Non-vascular	39,61
<i>X. citri</i> pv. <i>citri</i> 306	<i>Citrus</i> species	Citrus Canker	Non-vascular	62,63

Supplementary Table 3. Continued.

Specie	Host	Disease	Local of infection	Reference
<i>X. citri</i> pv. <i>malvacearum</i> AR81009	Cotton (<i>Gossypium hirsutum</i>)	Bacterial blight	Non-vascular	64,65
<i>X. citri</i> pv. <i>punicae</i> LMG7439	Pomegranate (<i>Punica granatum</i> L.)	Bacterial blight	Non-vascular	66,67
<i>X. citri</i> pv. <i>glycines</i> strain 12-2	Soybean (<i>Glycine max</i>)	Bacterial pustule	Non-vascular	68
<i>X. axonopodis</i> pv. <i>vasculorum</i> NCPPB 796	Sugarcane, Corn	Gumming disease/Bacterial leaf streak	Unknown	69,70
<i>X. vasicola</i> NCPPB 1060	Sorghum (<i>Sorghum vulgare</i>)	Bacterial leaf streak	Unknown	39,70
<i>X. vasicola</i> pv. <i>vasculorum</i> Xv1601	Corn (<i>Zea mays</i>)	Bacterial leaf streak	Unknown	39,71
<i>X. campestris</i> pv. <i>musacearum</i> NCPPB4379	Banana (<i>Musa</i> species)	Enset wilt	Vascular	72
<i>X. vasicola</i> NCPPB 902	<i>Tripsacum laxum</i>	Unknown	Unknown	70
<i>X. vasicola</i> pv. <i>arecae</i> NCPPB 2649	Betel palm (<i>Areca catechu</i>)	Unknown	Unknown	70
<i>X. oryzae</i> BAI23	Weed	Unknown	Unknown	73,74
<i>X. oryzae</i> pv. <i>oryzae</i> PXO99A	Rice (<i>Oryza sativa</i> L.)	Bacterial blight	Vascular	39,75
<i>X. oryzae</i> pv. <i>oryzicola</i> BLS256	Rice (<i>Oryza sativa</i> L.)	Bacterial leaf streak	Non-vascular	75,76
<i>X. translucens</i> pv. <i>translucens</i> DSM 18974	Barley (<i>Hordeum vulgare</i>)	Black chaff	Vascular	39,77,78
<i>X. translucens</i> pv. <i>undulosa</i> P3	Wheat (<i>Triticum</i> sp.)	Black chaff	Non-vascular	78–80
<i>Xanthomonas hyacinthi</i> CFBP 1156	Jacinth (<i>Hyacinthus</i>)	Yellow disease	Non-vascular	39,81
<i>X. sacchari</i> R1	Isolated from rice seeds	Nonpathogenic		82,83
<i>X. albilineans</i> Xa-FJ1	Sugarcane	Leaf scald	Vascular	84
<i>X. albilineans</i> GPE PC73	Sugarcane	Leaf scald	Vascular	84

Supplementary Table 4. Data collection and data refinement statistics for crystallographic data. Values in parenthesis represent the higher resolution shell. Carbohydrate validations were performed by Privateer software⁸⁵ (see details in Supplementary Table 5).

	XccXeg74 (GH74)	XacXaeA (CE20)	XacGalD (GH35)	XacGalD (GH35)
Ligand	XG	-	-	Galactose
PDB Code	7KN8	7KMM	7KMN	7KMO
Data collection				
Space group	P 21 21 21	P 1 21 1	I 4	I 4
Cell dimensions				
<i>a, b, c</i> (Å)	81.79, 100.43, 168.27	77.85, 63.50, 88.28	115.82, 115.82, 97.22	116.68, 116.68, 97.77
α, β, γ (°)	90.00	90.00, 98.15, 90.00	90.00	90.00
Resolution (Å)	38.80 - 1.95 (2.07 - 1.95)	87.39 - 1.90 (2.01 - 1.90)	19.86 - 1.80 (1.91 - 1.80)	46.03 - 1.75 (1.85 - 1.75)
<i>R</i> _{meas}	0.20 (1.09)	0.14 (1.55)	0.11 (2.13)	0.16 (2.94)
<i>I</i> / $\sigma(I)$	11.02 (2.50)	12.01 (1.36)	12.80 (0.85)	8.92 (0.64)
CC _{1/2} (%)	99.7 (88.8)	99.8 (61.7)	99.9 (37.6)	99.8 (32.7)
Completeness (%)	98.1 (93.7)	95.6 (94.6)	100.0 (99.9)	97.4 (99.8)
Redundancy	12.19 (10.12)	7.13 (7.04)	6.58 (6.07)	6.47 (5.96)
Refinement				
Resolution (Å)	19.99 - 1.95	19.97 - 1.90	19.86 - 1.80	19.45 - 1.75
No. reflections	99535	62206	59288	60752
<i>R</i> _{work} / <i>R</i> _{free}	0.216/0.252	0.190/0.238	0.177/0.210	0.193/0.242
No. atoms				
Protein	10852	6607	4040	4040
Ligand/ion	100/16	0/13	0/20	42/10
Water	1084	381	520	506
<i>B</i> -factors (Å ²)				
Protein	19.29	26.43	31.47	37.10
Ligand/ion	20.56/35.47	0/54.02	0/62.94	46.68/100.25
Water	20.28	28.59	36.83	41.88
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.007	0.007
Bond angles (°)	0.876	0.88	0.880	0.896
Carbohydrate validation				
Stereochemical problems	0	-	-	0
Unphysical puckering	0	-	-	0
Unlikely conformations	0	-	-	0

Supplementary Table 4. Continued.

	<i>XacXyl31</i> (GH31)	<i>XacXyl31</i> (GH31)	<i>XacAfc95</i> (GH95)
Ligand	-	Xylose	-
PDB Code	7KMP	7KNC	7KMQ
Data collection			
Space group	P 21 21 2	P 21 21 2	P 21 21 21
Cell dimensions			
a, b, c (Å)	117.10, 145.69, 62.87	117.49, 146.63, 63.04	103.91, 103.91, 173.75
α, β, γ (°)	90.00	90.00	90.00
Resolution (Å)	45.81 - 1.56 (1.64 - 1.56)	38.63 - 1.86 (1.97 - 1.86)	20.00 – 2.05 (2.17 – 2.05)
R_{meas}	0.07 (0.28)	0.09 (0.35)	0.08 (1.05)
$I / \sigma(I)$	20.15 (6.96)	17.65 (6.66)	13.15 (1.31)
$CC_{1/2}$ (%)	99.9 (97.9)	99.8 (96.2)	99.8 (46.5)
Completeness (%)	99.7 (98.7)	98.3 (91.2)	96.8 (96.1)
Redundancy	9.71 (9.64)	9.70 (9.41)	3.50 (3.22)
Refinement			
Resolution (Å)	45.81 - 1.56	38.63 - 1.87	19.78 – 2.05
No. reflections	155561	88980	117432
$R_{\text{work}} / R_{\text{free}}$	0.163/0.185	0.191/0.225	0.175/0.191
No. atoms			
Protein	7414	7348	11750
Ligand/ion	78/1	16/1	36/2
Water	1132	797	202
B -factors (Å ²)			
Protein	20.09	24.42	37.90
Ligand/ion	30.43/25.55	31.04/29.52	40.79/45.24
Water	31.73	30.17	36.27
R.m.s. deviations			
Bond lengths (Å)	0.007	0.004	0.002
Bond angles (°)	0.850	0.713	0.591
Carbohydrate validation			
Stereochemical problems	-	0	-
Unphysical puckering	-	0	-
Unlikely conformations	-	0	-

Supplementary Table 5. Privateer⁸⁵ analysis of the carbohydrates complexed in the crystallographic structures. Detailed monosaccharide data for *XccXeg74* (7KN8), *XacGalD* (7KMO) and *XacXyl31* (7KNC) carbohydrates complexes (Supplementary Table 4). The chain identifiers represent each ligand as defined in the respective PDB file. Names refer to the three-letter code correspondent to the monosaccharide in the model: BGC = β -glucose, YYS = α -xylose, GLA = α -galactose. The puckering amplitude Q is represented in Angstroms, and the angles φ and θ are represented in degrees. Puckering parameters were used to determine the conformations for pyranoses, as defined by Cremer-Pople⁸⁶. Anomer is the α/β geometric variation of the epimer, considering the anomeric carbon in the pyranoses. D/L is the isomeric form of the carbohydrate handedness found in the structure (D = right-handed, L = left-handed, N = unable to determine based solely in the structure). Conformation refers to the overall tridimensional structure adopted by the monosaccharide. RSCC is short for Real Space Correlation Coefficient, which measures the agreement between model and positive omit density. A RSCC below 0.8 is typically considered poor. *B*-factor column is the corresponding value for the monosaccharide, considering the overall crystallographic structure. Finally, the Diagnostic column indicates whether the monosaccharide is in a stereochemical and conformationally acceptable configuration, as evaluated by Privateer.

<i>XccXeg74</i> complexed with xyloglucan oligosaccharide (7KN8)										
Chain	Name	Q	φ	θ	Conformation	Anomer	D/L	RSCC	< <i>B</i> -factor>	Diagnostic
C	BGC	0.551	50.2387	15.6044	4C_1	β	D	0.80	22.4058	OK
C	BGC	0.535	224.838	2.65641	4C_1	β	D	0.89	19.0155	OK
C	YYS	0.530	109.429	7.7146	4C_1	α	N	0.84	22.1111	OK
D	BGC	0.529	161.316	7.70464	4C_1	β	D	0.84	21.8817	OK
D	BGC	0.552	269.654	4.8517	4C_1	β	D	0.89	19.2345	OK
D	YYS	0.550	89.8472	12.1197	4C_1	α	N	0.81	21.6567	OK
<i>XacGalD</i> complexed with galactose (7KMO)										
Chain	Name	Q	φ	θ	Conformation	Anomer	D/L	RSCC	< <i>B</i> -factor>	Diagnostic
C	GLA	0.552	261.192	8.60379	4C_1	α	D	0.79	35.3717	Ok
<i>XacXyl31</i> complexed with xylose (7KNC)										
Chain	Name	Q	φ	θ	Conformation	Anomer	D/L	RSCC	< <i>B</i> -factor>	Diagnostic
C	YYS	0.546	121.95	8.0779	4C_1	α	N	0.82	27.137	Ok

[illegible]

[illegible]

Supplementary Table 8. Biochemical parameters of XyGUL and accessory enzymes. Kinetic parameters were obtained in the preferred conditions. pNP = *para*-nitrophenyl, Galp = galactopyranoside, Xylp = xylopyranoside, Fucp = fucopyranoside and Glup = glucopyranoside.

Enzyme	Substrate	pH	T (°C)	K _{0.5} (mmol.L ⁻¹ or mg.mL ⁻¹ *)	k _{cat} (s ⁻¹)	k _{cat} /K _{0.5} (L.mmol ⁻¹ .s ⁻¹ or L.g ⁻¹ .s ⁻¹ *)
<i>XacXeg74</i>	Tamarind xyloglucan	6.0	35	3.34 ± 0.58*	4.40 ± 0.26	1.32*
<i>XacXaeA</i>	pNP-acetate	7.5	20	0.66 ± 0.05	1.59x10 ⁴ ± 244	2.41.10 ⁴
<i>XacGalD</i>	pNP-β-D-Galp	5.0	50	1.06 ± 0.09	1.42x10 ⁶ ± 21807	1.34.10 ⁶
<i>XacXyl31</i>	pNP-α-D-Xylp	6.0	45	1.65 ± 0.19	1.29x10 ⁵ ± 3997	7.81.10 ⁴
<i>XacAfc95</i>	pNP-α-L-Fucp	6.0	55	3.24 ± 0.37	3.13x10 ⁴ ± 1757	9.66.10 ³
	pNP-β-D-Glup			5.23 ± 0.19	141.55 ± 4.42	27.07
<i>XacBgl3A</i>	β-1,4-glucobiose	5.5	45	38.72 ± 24.03*	5.65 ± 1.32	0.15*
	Laminarin			2.30 ± 0.07*	3.01 ± 0.08	1.31*
	pNP-β-D-Glup			3.68 ± 0.25	32.83 ± 0.54	8.92
<i>XacBgl3B</i>	β-1,4-glucobiose	6.0	25	1.35 ± 0.15*	8.16 ± 0.19	6.04*
	Laminarin			5.93 ± 0.20*	2.47 ± 0.06	0.42*
	pNP-β-D-Glup			3.36 ± 0.19	13.46 ± 0.32	4.01
<i>XacBgl3C</i>	β-1,4-glucobiose	7.5	35	16.08 ± 1.10*	3.73 ± 0.09	0.23*
	Laminarin			11.39 ± 0.26*	1.73 ± 0.03	0.15*
<i>XacEgl9</i>	Tamarind xyloglucan	6.0	35	4.82 ± 0.11*	4.04 ± 0.07	0.84*

Supplementary Table 9. Composition, structure and nominal masses of ions generated by ESI(+)-MS of xyloglucan oligosaccharides obtained from *Arabidopsis thaliana* cell wall, as depicted in the supplementary Fig. 8. Letters indicate the type of substitutions appended to the glucose backbone of the identified oligosaccharides. G = non-substituted glucose, X = glucose substituted with a xylose at C-6, L = X with a galactose appended at xylose C-2 and F = L with a fucose appended at galactose C-2. OAc indicates the *O*-acetyl substituent at galactose C-6. Hex = hexose, Pen = pentose and Dox = deoxyhexose. M = nominal mass.

Nominal mass	Composition	Suggested structure
1247	Hex ₅ Pen ₃ [M + Na] ⁺	XXLG/XLXG
1289	Hex ₅ Pen ₃ OAc ₁ [M + Na] ⁺	XXLG/XLXG + 1OAc
1393	Hex ₅ Pen ₃ Dox ₁ [M + Na] ⁺	XXFG
1435	Hex ₅ Pen ₃ Dox ₁ OAc ₁ [M + Na] ⁺	XXFG + 1OAc
1555	Hex ₆ Pen ₃ Dox ₁ [M + Na] ⁺	XLFG
1597	Hex ₆ Pen ₃ Dox ₁ OAc ₁ [M + Na] ⁺	XLFG + 1OAc

Supplementary Table 10. SAXS analysis of XyGUL enzymes. The molecular weight calculated from the data by SAXSMoW⁸⁸ server (MW_{MoW}) was compared to the predicted molecular weight derived from the amino acid sequence (MW_{AA}). Maximum diameter (D_{max}), radius of gyration (R_g), molecular volume (V_m) and Porod volume (V_P) were calculated by Primusqt software⁸⁹. If it applies, oligomeric interface interaction energy was also calculated by PDBePISA server⁹⁰. Negative values in this field represent favorable quaternary structure assemblies. *XacGalD* was analyzed by SEC-MALS experiments and the results are not listed in this table (see Supplementary Fig. 17).

Enzyme	D_{max} (nm)	R_g (nm)	V_m (nm ³)	V_P (nm ³)	MW_{MoW} (kDa)	MW_{AA} (kDa)	Δ^iG (kcal.mol ⁻¹)
<i>XacXeg74</i>	9.00	3.16	98.20	97.11	81.02	79.44	-
<i>XacXaeA</i>	10.60	3.28	87.32	92.09	72.04	68.48	-
<i>XacGalD</i>	-	-	-	-	-	59.58	-4.4
<i>XacXyl31</i>	9.80	3.61	121.28	127.30	100.06	118.4	-
<i>XacAfc95</i>	14.8	4.09	227.05	226.03	187.32	86.68	-5.7

Supplementary Table 11. GH35 structures reported in the CAZy database. The table lists at least one representative structure (when available) for each organism in the GH35 family (PDB entry column), sorted by Domain and Phylum. The quaternary structure types were determined using the PDBePISA server⁹⁰ (See Supplementary Fig. 20 for the examples highlighted in bold in the table), with each respective assembly solvation free energy gain presented as ΔG^{int} (kcal.mol⁻¹). A negative value in this column means a favorable calculated quaternary structure. The D562 residue from the other subunit that interacts with the galactose at the -1 subsite is indicated when conserved. The experimentally determined activity is also listed as annotated in the CAZy database.

Domain	Organism	PDB entry	Quaternary structure	ΔG^{int} (PISA)	Accessory domains	D562 conservation	Function	Reference
Archaea (Thermococci)	<i>Pyrococcus furiosus</i> DSM 3638	6JOW	Dimer (I)	-59.8	2	-	-	-
Archaea (Thermococci)	<i>Pyrococcus horikoshii</i> OT3	5GSL	Dimer (I)	-31.4	2	-	exo- β -glucosaminidase (chitosan)	-
Archaea (Thermococci)	<i>Thermococcus kodakarensis</i> KOD1	5GSM	Dimer (I)	-79	2	-	exo- β -glucosaminidase (chitosan)	-
Bacteria (Firmicutes)	<i>Bacillus circulans</i> ATCC 31382	4MAD	Dimer (II)	-2.7	2	-	β -galactosidase (β -1,3)	15
Bacteria (Firmicutes)	<i>Streptococcus pneumoniae</i> TIGR4	4E8D	Dimer (II)	-10.9	2	-	β -galactosidase (β -1,3)	91
Bacteria (Bacteroidetes)	<i>Bacteroides thetaiotaomicron</i> VPI-5482	6EON	Monomer (I)	-	3	-	-	17
Bacteria (Proteobacteria)	<i>Caulobacter vibrioides</i> CB15	3U7V	Tetramer	-26.7	1	D527	β -galactosidase	-
Bacteria (Proteobacteria)	<i>Cellvibrio japonicus</i> Ueda107	5JAW	Tetramer	-156.4	1	D550	β -galactosidase (Xyloglucan)	92
Bacteria (Proteobacteria)	<i>Xanthomonas citri</i> pv. citri 306	7KMN	Tetramer	-284.9	1	D562	β -galactosidase (Xyloglucan)	This work
Eukaryota (Fungi)	<i>Aspergillus oryzae</i> RIB40	4IUG	Monomer (II)	-	4	-	β -galactosidase (lactose, β -1,3 and β -1,4)	18
Eukaryota (Fungi)	<i>Penicillium</i> sp.	1TG7	Monomer (II)	-	4	-	β -galactosidase	93
Eukaryota (Fungi)	<i>Trichoderma reesei</i>	3OG2	Monomer (II)	-	4	-	β -galactosidase	94
Eukaryota (Viridiplantae)	<i>Solanum lycopersicum</i>	3W5F	Monomer (II)	-	4	-	β -galactosidase (β -1,3 and β -1,4, pectin)	-
Eukaryota (Metazoa)	<i>Homo sapiens</i>	3THC	Dimer (III)	-41.4	2	-	β -galactosidase	16

Supplementary Table 12. Expression profile of GH3 β -glucosidase genes from *X. citri* in RNA-seq assays. Data are shown as mean values of TPM \pm SD from four independent biological replicates. TPM: Transcripts *Per Million* reads.

<i>Locus</i>	Protein	TPM \pm SD
XAC1448	<i>XacBgl3A</i>	2.8 \pm 0.8
XAC1793	<i>XacBgl3B</i>	15.3 \pm 4.6
XAC3869	<i>XacBgl3C</i>	35.5 \pm 19.4

Supplementary Table 13. Differentially expressed genes in RNA-seq assays corresponding to the type III effector proteins (T3E) from *X. citri* pv. *citri* 306. Genes were considered differentially expressed according to the Wald test implemented in the DESeq2 package. *p*-values were adjusted for multiple tests using Benjamini-Hochberg (BH) method also implemented in DESeq2 package. Thresholds: $|\log_2 \text{fold change}| \geq 1$ and *p* adjusted ≤ 0.05 . The overview of T3E can be assessed in *Xanthomonas* Resource (<http://www.xanthomonas.org/t3e.html>).

<i>Locus</i>	\log_2 fold change	<i>p</i> -adjusted	Description
XAC0277	3.03	1.75E-17	XopR
XAC0286	3.59	2.41E-34	XopE1
XAC0754	2.76	5.74E-11	XopI
XAC1171	3.81	4.65E-19	XopAU
XAC1172	3.36	3.00E-19	XopAV
XAC1208	1.64	4.76E-02	XopP
XAC2009	2.23	3.68E-07	XopZ1
XAC2786	2.70	1.37E-16	XopN
XAC2922	3.67	2.28E-28	HrpW
XAC3085	2.67	2.77E-09	XopK
XAC3090	2.95	2.03E-13	XopL
XAC3666	2.54	1.18E-10	XopAK
XAC4213	2.11	4.65E-19	XopAD
XAC4333	3.03	4.17E-11	XopQ

Supplementary Table 14. Putative major facilitator superfamily (MFS) transporters from *X. citri* pv. *citri* 306. Corresponding genes were identified by search in the Pfam database (<https://pfam.xfam.org/>).

<i>Locus</i>	<i>Gene</i>	<i>Description</i>
XAC0110	<i>proP</i>	proline-betaine transporter
XAC0229		MFS transporter
XAC0303	<i>opdE</i>	transcriptional regulator
XAC0317	<i>ynfM</i>	MFS transporter
XAC0349	<i>vanK</i>	MFS transporter
XAC0507	<i>aas</i>	2-Acylglycerophosphoethanolamine acyltransferase
XAC0509		MFS transporter
XAC0642	<i>rmrB</i>	MFS transporter
XAC0712	<i>gluP</i>	glucose-galactose transporter
XAC1215	<i>bcr</i>	MFS transporter
XAC1363	<i>araJ</i>	MFS transporter
XAC1446	<i>pmrB</i>	multidrug resistance membrane translocase
XAC1450	<i>ygdR</i>	oligopeptide transporter
XAC1556	<i>fucP</i>	glucose-galactose transporter
XAC1705		MFS transporter
XAC1777	<i>xylE</i>	MFS transporter
XAC1801	<i>proP</i>	Prop transport protein
XAC2161	<i>tetV</i>	MFS transporter
XAC2234	<i>cynX</i>	MFS transporter
XAC2340	<i>cynX</i>	MFS transporter
XAC2356	<i>tetV</i>	drug/proton antiporter
XAC2474	<i>rmrB</i>	transport protein
XAC2484	<i>yhjE</i>	metabolite transport protein
XAC2488	<i>yhjX</i>	integral membrane transporter
XAC2494	<i>yieO</i>	drug resistance translocase
XAC2597	<i>sucI</i>	transport protein
XAC2837	<i>araJ</i>	MFS transporter
XAC3001	<i>ptr</i>	MFS transporter
XAC3027	<i>emrA</i>	MFS transporter

Supplementary Table 14. (Continued). Putative major facilitator superfamily (MFS) transporters from *X. citri* pv. *citri* 306. Corresponding genes were identified by search in the Pfam database (<https://pfam.xfam.org/>).

<i>Locus</i>	<i>Gene</i>	<i>Description</i>
XAC3056		conserved hypothetical protein
XAC3157	<i>ycaD</i>	transmembrane transport protein
XAC3179	<i>yceE</i>	transport protein
XAC3474	<i>citI</i>	citrate carrier protein
XAC3488	<i>sucI</i>	sugar transporter
XAC3901	<i>ampG</i>	signal transducer
XAC4190	<i>fucP</i>	fucose permease
XAC4196	<i>ynaJ</i>	cation symporter
XAC4255	<i>exuT</i>	hexuranate transporter
XAC4295	<i>tetA</i>	tetracycline-efflux transporter
XAC4308	<i>kgtP</i>	dicarboxylate transport protein
XAC4361	<i>ttuB</i>	MFS transporter

Supplementary Table 15. Specific activity of endoglucanases from *X. citri* on XyG. Tamarind xyloglucan (2.5 mg.mL⁻¹) was used for the assay in conditions of pH and temperature detailed in the Supplementary Table 21. All constructs lack the N-terminal signal peptide. ^a catalytic domain. ^b Deletion of 27 residues of high hydrophobicity at C-terminus. N.D. = not detected. ^c ⁹⁵ (V_0 =initial velocity and $[E]_t$ =enzyme concentration).

Gene	Protein	Construct (a.a.)	Family	$V_0/[E]_t$ (s ⁻¹)
XAC2522	<i>XacEgl9</i>	(22-586)	GH9	1.00 ± 0.03
XAC0029	<i>XacEgl5B</i>	(27-350)	GH5_5	0.34 ± 0.03
XAC0612	<i>XacEngXCA</i>	(26-368) ^a	GH5_1	0.10 ± 0.07
XAC0030	<i>XacEgl5C</i>	(37-357)	GH5_5	N.D.
XAC0028	<i>XacEgl5A</i>	(31-350) ^b	GH5_5	N.D.
XAC0346	-	(53-453)	GH5	N.D.
XAC3516	<i>XacCel8</i>	(26-384)	GH8	N.D. ^c

Supplementary Table 16. List of primers and vectors used in gene cloning for the heterologous expression of XyGUL and accessory enzymes. Restriction sites are underlined. pET28a-XAC0346, pET28a-XAC1275 and pET28a-XAC3869 clones were purchased from GenScript.

Protein	Vector	Primer sequence (5'→3')	Restriction site
<i>XacBgl3A</i>	pET28a	F: GGAATTCCATATGCAGGGCGCGCCATCTTCGC	NdeI
		R: CCGCTCGAGTTACGGCAACTGTGCCGC	XhoI
<i>XccXeg74</i>	pETM11	F: CAGGGCGCCATGGCCACGTCCGGGC	-
		R: GACCCGACGCGGTTATCTCGGATCGCCGTAG	-
<i>XacXeg74</i>	pET28a	F: CATATGGATGAGCCAGGTACGCCA	NdeI
		R: AAGCTTTTATCGTGGATCGCCATAGAAGAT	HindIII
<i>XacXaeA</i>	pET28a	F: CATATGGTGCCAACACTACCGCTG	NdeI
		R: AAGCTTTTACCAGGTATCGGTGCG	HindIII
<i>XacGalD</i>	pET28a	F: CATATGCAGACGCCGATGCCGCA	NdeI
		R: AAGCTTTTATTTGTAGGTTGCCAGTTTGATCTTGAGCAG	HindIII
<i>XacXyl31</i>	pET28a	F: CATATGCAGGAAGTGCGCAAGGC	NdeI
		R: CTCGAGTACGCTTTGCCCCGACGCAATC	XhoI
<i>XacAfc95</i>	pET28a	F: GAATTCCATATGCAGGCAACGCAGGCTTCGAA	NdeI
		R: GTCGACTTATTGCGTCACCAATCGGT	SalI
<i>XacBgl3B</i>	pET28a	F: GCGGCCGCACATATGGGCAAGGACACTGCTGCC	NdeI
		R: CTCGAGTTACTTGACAGGGCAGTCGACGGT	XhoI
<i>XacEgl9</i>	pET28a	F: CATATGGCCGAGACGCCTGGCAGG	NdeI
		R: CTCGAGTTAGCGAGTCGACGCCTCGAT	XhoI
<i>XacAbf43A</i>	pET28a	F: ATATATGCTAGCATGGCGACGCAGCGCGCGGG	NheI
		R: TATATAGGATCCTTACGTCAGCGCGCGATAGC	BamHI
<i>XacEgl5A</i>	pET28a	F: CATATGACGCGCGCTCAC	NdeI
		R: CTCGAGTTAACGTGTGGCGAGTGTG	XhoI
<i>XacEgl5B</i>	pET28a	F: CATATGCAAAGCGCCACCGGCCTGAAG	NdeI
		R: CTCGAGTTAATCGGTAATCCGGCGCGCA	XhoI
<i>XacEgl5C</i>	pET28a	F: CATATGCTGAAGTATGTTGGCGTCAATC	NdeI
		R: GTCGACTTATGCGTACTTGCTAGGATC	SalI
<i>XacEngXCA</i>	pET28a	F: CATATGATGTATTCGGTCAGCAATAA	NdeI
		R: AAGCTTTCACCACAGCGTCCGCAA	HindIII

Supplementary Table 17. List of primers used for site-directed mutagenesis of XyGUL enzymes.

Protein	Residue change	Primer sequence (5'→3')
<i>XacGalD</i>	D562A	F: CCGTAATCGGTCTGGGCGCCATTCCAGTTGC
		R: GCAACTGGAATGGCGCCCAGACCGATTACGG
	S106R	F: CCGGGATAGTTGCTGCGGTTGTTGACCTGCGC
		R: GCGCAGGTCAACAACCGCAGCAACTATCCCGG
<i>XacXyl31</i>	W328A	F: TGATTCCACGGATTCGCGTTCTGGCGCCAGCG
		R: CGCTGGCGCCAGAACGCGAATCCGTGGAATCA
<i>XacAfc95</i>	T395H	F: CTCGGTGTTGATGTTGATGTGGTACTTGCTTTCCCACGGC
		R: GCCGTGGGAAAGCAAGTACCACATCAACATCAACACCGAG

Supplementary Table 18. Protein expression conditions.

Protein	Strain	Temperature	Time	Medium	[IPTG]
XAC1275	BL21 (DE3) pRARE II	18 °C	16 h	Auto-induction	-
<i>XacBgl3A</i>	BL21(DE3)	20°C	20 h	TB	0.5 mM
<i>XccXeg74</i>	Rosetta 2 (DE3) pLysS	20°C	25 h	Auto-induction	-
<i>XacXeg74</i>	BL21(DE3) pLysS	20°C	16 h	LB	0.5 mM
<i>XacXaeA</i>	BL21(DE3) pLysS	30 °C	4 h	LB	0.5 mM
<i>XacGalD</i>	BL21(DE3) pLysS	30 °C	4h	LB	0.5 mM
<i>XacXyl31</i>	BL21(DE3) pLysS	20°C	16 h	LB	0.5 mM
<i>XacAfc95</i>	BL21(DE3) pLysS	20°C	16 h	LB	0.5 mM
<i>XacBgl3B</i>	BL21(DE3)	20°C	20 h	TB	0.5 mM
<i>XacEgl9</i>	BL21(DE3)	20°C	16 h	TB	0.2 mM
<i>XacBgl3C</i>	BL21(DE3)	20°C	20 h	TB	0.5 mM
<i>XacAbf43A</i>	BL21(DE3)	18 °C	16 h	TB	0.5 mM
<i>XacEgl5A</i>	BL21(DE3) SHuffle	20 °C	16 h	TB	0.2 mM
<i>XacEgl5B</i>	BL21(DE3)	20 °C	16 h	LB	0.1 mM
<i>XacEgl5C</i>	BL21(DE3)	20 °C	16 h	LB	0.1 mM
<i>XacEngXCA</i>	BL21(DE3) Δ slyD pRARE II	20 °C	16 h	Auto-induction	-
XAC0346	BL21(DE3)	20°C	20 h	TB	0.5 mM

Supplementary Table 19. Protein purification conditions.

Protein	Lysis buffer	Affinity buffer	Gradient of imidazole	Size-exclusion buffer
<i>XacBgl3A</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mM NaCl, 20 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF, and 0.1 mg. mL ⁻¹ lysozyme	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mmol.L ⁻¹ NaCl	20-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XccXeg74^a</i>	20 mmol.L ⁻¹ Tris-HCl pH 8.0, 300 mmol.L ⁻¹ NaCl, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ Tris-HCl pH 8.0, 150 mmol.L ⁻¹ NaCl,	50-300 mmol.L ⁻¹	20 mmol.L ⁻¹ Tris-HCl pH 8.0, 150 mmol.L ⁻¹ NaCl
<i>XacXeg74</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 5 mmol.L ⁻¹ imidazole, 150 mmol.L ⁻¹ NaCl, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl	5-1000 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium citrate pH 5.5, 150 mmol.L ⁻¹ NaCl
<i>XacXaeA^b</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 5 mmol.L ⁻¹ imidazole, 150 mmol.L ⁻¹ NaCl	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl	5-1000 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium acetate 5.5, 150 mmol.L ⁻¹ NaCl
<i>XacGalD</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 5 mmol.L ⁻¹ imidazole, 150 mmol.L ⁻¹ NaCl, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl	5-1000 mmol.L ⁻¹	20 mmol.L ⁻¹ Hepes pH 7.5, 150 mmol.L ⁻¹ NaCl
<i>XacXyl31</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 5 mmol.L ⁻¹ imidazole, 150 mmol.L ⁻¹ NaCl, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl	5-1000 mmol.L ⁻¹	20 mmol.L ⁻¹ Hepes pH 7.5, 150 mmol.L ⁻¹ NaCl
<i>XacAfc95</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 5 mmol.L ⁻¹ imidazole, 150 mmol.L ⁻¹ NaCl, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl	5-1000 mmol.L ⁻¹	20 mmol.L ⁻¹ Hepes pH 7.5, 150 mmol.L ⁻¹ NaCl
<i>XacBgl3B</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mmol.L ⁻¹ NaCl, 20 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF, and 0.1 mg. mL ⁻¹ lysozyme	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mmol.L ⁻¹ NaCl	20-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XacEgl9</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 700 mmol.L ⁻¹ NaCl, 5 mmol.L ⁻¹ imidazole, 5 % glycerol, pH 7.4, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 700 mmol.L ⁻¹ NaCl, 5% glycerol	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ MES, pH 6.0, 700 mmol.L ⁻¹ NaCl

^a*XccXeg74* was subjected to removal of the 6xHis-tag by TEV protease after affinity purification and dialysis to 20 mmol.L⁻¹ Tris-HCl at pH 8.0, 300 mmol.L⁻¹ NaCl buffer solution. TEV protease reaction was carried for 24 hours at 8 °C. The protein solution was submitted to size-exclusion chromatography.

^b*XacXaeA* was subjected to limited proteolysis after nickel chromatography using 0.01% (*m/m*) trypsin for 4 hours at 4°C. Proteolysis reaction was stopped by adding 1 mmol.L⁻¹ phenylmethylsulfonyl fluoride (PMSF) and the resulting sample was purified by anion exchange chromatography using NaCl linear gradient from 100 mmol.L⁻¹ to 1 mol.L⁻¹ in a HiTrapQ HP column coupled to an ÄKTA purifier system, using 50 mmol.L⁻¹ acetate buffer at pH 5.5. Fractions were analyzed by SDS-PAGE and submitted to size-exclusion chromatography.

Supplementary Table 19. Continued.

Protein	Lysis buffer	Affinity buffer	Gradient of imidazole	Size-exclusion buffer
<i>XacBgl3C</i>	20 mmol.L ⁻¹ sodium phosphate, pH 7.4, 300 mmol.L ⁻¹ NaCl, 20 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF, and 0.1 mg.mL ⁻¹ lysozyme	20 mmol.L ⁻¹ sodium phosphate, pH 7.4, 300 mmol.L ⁻¹ NaCl	20-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate buffer, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XacAbf43A</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl, 5% glycerol, 5 mmol.L ⁻¹ imidazole, 4 mmol.L ⁻¹ PMSF, 2 mmol.L ⁻¹ benzamidine and 0.1 mg.mL ⁻¹ lysozyme	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 150 mmol.L ⁻¹ NaCl and 5% glycerol	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ Hepes pH 7.5, 150 mmol.L ⁻¹ NaCl and 10% glycerol
<i>XacEgl5A</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl, 5 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl and 5% glycerol	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate buffer, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XacEgl5B</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl, 5 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate buffer, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XacEgl5C</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl, 5 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate buffer, pH 7.4 and 150 mmol.L ⁻¹ NaCl
<i>XacEngXCA</i>	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl, 5 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 500 mmol.L ⁻¹ NaCl	5-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate buffer, pH 7.4 and 150 mmol.L ⁻¹ NaCl
XAC0346	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mmol.L ⁻¹ NaCl, 20 mmol.L ⁻¹ imidazole, 1 mmol.L ⁻¹ PMSF, and 0.1 mg. mL ⁻¹ lysozyme	20 mmol.L ⁻¹ sodium phosphate pH 7.4, 300 mmol.L ⁻¹ NaCl	20-500 mmol.L ⁻¹	20 mmol.L ⁻¹ sodium phosphate, pH 7.4 and 150 mmol.L ⁻¹ NaCl

Supplementary Table 20. Crystallization conditions. Proteins were concentrated in the final purification buffer (Supplementary Table 19). Ligand introduced by cocrystallization were incubated for 1 h at 4 °C under mild agitation and centrifuged for 15 min (4 °C) at 14,000 g prior to the experiment. Crystallization experiments were performed at 18 °C in a temperature-controlled environment. A 1:1 drop ratio of protein to crystallization solution was employed. All proteins were crystallized by the vapor diffusion technique in hanging or sitting drops.

Enzyme	Enzyme concentration ($\mu\text{mol.L}^{-1}$)	Ligand	Condition	Method
<i>XccGH74</i>	245	XXG ¹ Cocrystallization 5 mmol.L ⁻¹	0.2 mol.L ⁻¹ sodium iodide 20% (v/v) PEG 3350 0.1 mol.L ⁻¹ bis-tris propane pH 7.5 0.01 mol.L ⁻¹ zinc chloride	Sitting drop
<i>XacXaeA</i>	160	-	20% (v/v) PEG 6000 0.1 mol.L ⁻¹ tris pH 7.0 0.1 mol.L ⁻¹ lithium sulfate	Hanging drop
<i>XacGalD</i>	336	-	1.26 mol.L ⁻¹ ammonium sulfate 0.1 mol.L ⁻¹ tris pH 8.5	Sitting drop
<i>XacGalD</i>	235	Galactose ² Cocrystallization 5 mmol.L ⁻¹	1.26 mol.L ⁻¹ ammonium sulfate 8% (v/v) glycerol 0.1 mol.L ⁻¹ tris pH 8.5	Sitting drop
<i>XacXyl31</i>	312	-	0.2 mol.L ⁻¹ potassium nitrate 20% (v/v) PEG 3350	Sitting drop
<i>XacXyl31</i>	312	Xylose ² 4 h soaking 10 mmol.L ⁻¹	0.2 mol.L ⁻¹ potassium formate 20% (v/v) PEG 3350	Sitting drop
<i>XacAfc95</i>	185	-	20% (v/v) 2-methyl-2,4-pentanediol (MPD) 0.1 mol.L ⁻¹ tris pH 8.5	Sitting drop

¹Megazyme

²Sigma-Aldrich

Supplementary Table 21. Enzymatic reaction conditions of XyGUL and accessory enzymes used for activity detection screening or specific activity assays.

Protein	Enzyme concentration	Buffer	Reaction with polymeric substrates	Reaction with synthetic substrates
<i>XacBgl3A</i>	100 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 20 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacBgl3B</i>	100 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 20 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacBgl3C</i>	100 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 20 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacXeg74</i>	150 $\mu\text{g.mL}^{-1}$	80 mmol.L^{-1} McIlvaine buffer pH 5.5	2 mg.mL^{-1} of substrate, 16 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacXaeA</i>	150 $\mu\text{g.mL}^{-1}$	80 mmol.L^{-1} HEPES buffer pH 7.5	2 mg.mL^{-1} of substrate, 16 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacGalD</i>	150 $\mu\text{g.mL}^{-1}$	80 mmol.L^{-1} McIlvaine buffer pH 5.5	2 mg.mL^{-1} of substrate, 16 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacXyl31</i>	150 $\mu\text{g.mL}^{-1}$	80 mmol.L^{-1} McIlvaine buffer pH 5.5	2 mg.mL^{-1} of substrate, 16 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacAfc95</i>	150 $\mu\text{g.mL}^{-1}$	80 mmol.L^{-1} McIlvaine buffer pH 5.5	2 mg.mL^{-1} of substrate, 16 hours at 35 °C	2 mmol.L^{-1} of substrate, 1 hour at 35 °C
<i>XacEgl9</i>	100 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} MES buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 10 min at 35 °C	-
<i>XacAbf43A</i>	80 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 6 hours at 37 °C	2.5 mmol.L^{-1} of substrate, 4 hours at 37 °C
<i>XacEgl5A</i>	20 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 6.0	2.5 mg.mL^{-1} of substrate, 10 min at 35 °C	-
<i>XacEgl5B</i>	10 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 5.0	2.5 mg.mL^{-1} of substrate, 10 min at 35 °C	-
<i>XacEgl5C</i>	60 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 5.0	2.5 mg.mL^{-1} of substrate, 10 min at 35 °C	-
<i>XacEngXCA</i>	6 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 5,5	2.5 mg.mL^{-1} of substrate, 10 min at 65 °C	-
XAC0346	200 $\mu\text{g.mL}^{-1}$	40 mmol.L^{-1} McIlvaine buffer pH 7.0	2.5 mg.mL^{-1} of substrate, 20 hours at 35 °C	-

Supplementary Table 22. Summary of RNA-seq data. Each sample represents a biologically independent experiment.

Sample	Input reads	QC reads	QC reads (%)	rRNA reads (%)	Mapped reads (%)
XVM2m_Glucose 1	2895851	2524189	87.17	5.64	87.68
XVM2m_Glucose 2	5660364	4990317	88.16	1.92	91.82
XVM2m_Glucose 3	5540215	5018731	90.59	2.58	68.56
XVM2m_Glucose 4	5164471	4599580	89.06	0.79	69.5
XVM2m_Glucose 5	4847539	4237597	87.42	14.36	61.53
XVM2m_Glucose 6	23704254	19279704	81.33	22.76	73.24
XVM2m_XyGOs 1	11228037	9610422	85.59	1.26	86.44
XVM2m_XyGOs 2	5107135	4483206	87.78	4.19	86.75
XVM2m_XyGOs 3	3738110	3113289	83.29	0.87	85.47
XVM2m_XyGOs 4	4970294	4324291	87.00	9.53	78.12

Supplementary Table 23. Classification of candidate reference genes based on the analysis of RNA-seq data under different culture conditions^a. The genes are classified from the lowest to the highest coefficient of variation. Values highlighted in bold indicate genes with normal distribution according to the Shapiro-Wilk test (p -value > 0.05). CV: coefficient of variation; MFC: ratio of the maximum to minimum TPM value of each gene. Genes with the highest mean TPM values and p -value > 0.05 were considered for further analysis, except for XAC1735 due to the presence of non-specific products in preliminary RT-PCR assays.

<i>Locus</i>	Description	Mean TPM	CV	MFC	<i>p</i> -value
XAC4132	acid phosphatase	1.1	0.13	1.48	2.80E-06
XAC1735	hfq	555.62	0.17	1.87	0.48
XAC2293	epimerase	137.66	0.18	1.9	0.16
XAC0145	conserved hypothetical protein	5.72	0.19	2.18	0.73
XAC0328	multidrug efflux transporter	5.05	0.19	2.41	0.29
XAC2491	conserved hypothetical protein	120.94	0.19	1.94	0.03
XAC2177	hypothetical protein	305.45	0.2	1.94	0.08
XAC4047	glutathione S-transferase	111.24	0.2	2.67	0.94
XAC4142	conserved hypothetical protein	1.32	0.2	2.01	0.01
XAC3440	H ⁺ translocating pyrophosphate synthase	34.53	0.2	2.29	0.33
XAC2594	threonyl-tRNA synthetase	291.65	0.2	2.44	0.44
XAC2416	virulence regulator	457.34	0.2	2.26	0.02
XAC3667	outer membrane protein	115.84	0.2	2.35	0.14
XAC3701	Na ⁺ :H ⁺ antiporter	37.23	0.2	2.68	0.53
XAC4140	ClpB	2.34	0.21	2.11	0.11
XAC4218	sec-independent protein translocase	584.53	0.21	2.47	0.14
XACb0065	avirulence protein	78.13	0.21	2.69	0.27
XAC3784	hypothetical protein	120.9	0.21	2.69	0.19
XAC0924	NAD(P) transhydrogenase subunidad β	52.1	0.21	2.16	0.41
XAC4116	serine/threonine kinase	2.12	0.21	2.48	0.77

^a Modified XVM2 medium containing glucose, β-1,4-glucobiose, starch, xyloglucan oligosaccharides, xylan oligosaccharides, arabinoxylan oligosaccharides or galactomannan oligosaccharides as sole carbohydrate source.

Supplementary Table 24. Analysis of gene expression stability of candidate reference genes. The variation in expression levels was evaluated for *X. citri* growth in minimal medium XVM2m containing different carbohydrate sources. Results were generated by BestKeeper⁹⁶, NormFinder⁹⁷, and RefFinder⁹⁸ algorithms. The ranking of each analysis is in parentheses, with 1 being considered the most stable and 4 being the least stable.

<i>Locus</i>	BestKeeper SD [\pm CP]	NormFinder Stability value	RefFinder Geometric mean
XAC4218	0.24 (1)	0.087 (1)	1.68 (2)
XAC2177	0.30 (2)	0.201 (2)	1.0 (1)
XAC4047	0.39 (3)	0.245 (3)	3.22 (3)
XAC2293	0.44 (4)	0.317 (4)	3.72 (4)

Supplementary Table 25. Primers used in RT-qPCR assays. Standard curves were generated from serially diluted samples (10X) to evaluate the efficiency of RT-qPCR reactions (E) and correlation coefficients (R^2). The reactions with a final volume of 10 μ L and with 2 ng of RNA were performed using the following protocol: 48 °C for 30 min (cDNA synthesis), followed by a step at 95 °C for 10 min, and then 40 cycles at 95 °C for 15 sec (denaturation), 60 °C for 30 sec (annealing) and 72 °C for 30 sec (extension). Primers corresponding to candidate reference genes are highlighted in bold^a.

<i>Locus</i>	Forward primer	Reverse primer	Concentration (μ M)	E ^a (%)	R ²	Amplicon length (pb)	Reference
XAC0394	AAGCCAGCTCCGAGATTAC	CTTCCTTGCCATCGCCGTAA	0.15	110	0.992	96	This study
XAC0397	ACGGGCAGATGAGTTCGTTGA	GCCACGTTGAAGTCCAGATCATT	0.10	97	0.996	109	This study
XAC0409	TCCTATTCGCTGGAGGTGCTA	TTCGGTGGGTGTCGAGATCA	0.20	101	0.999	105	This study
XAC0416	CGCCAACTCGTCCTTCTTTCAG	GGTCCAGTTGCTTTTCCGAGATG	0.10	102	0.999	95	This study
XAC1265	CACCTGAGCGTTGGTCCATA	AGCCAGGCAATCGAGAACT	0.15	105	0.999	104	This study
XAC1266	AGCGATCTCTGCGTTGTCCTAC	ATACGCATCTTCGGCCTCTTCCTGA	0.10	107	0.994	211	⁹⁹
XAC2293	CGGATCGTGCTGCTGGAAAT	GCCGCAATCGCCTAACACT	0.20	104	0.994	108	This study
XAC4047	CACCCTCAAGCCGGTCAAC	CGGTGCATGGTCGACGAT	0.10	101	0.999	103	This study
XAC4218	ATCGTGCTGGTGATCGTGTT	TCGTCGTGCATGCCTTTCTT	0.10	104	0.995	107	This study
XAC2177	GGTGTAGCCGTCTTTGTCTGC	CAGAACATCGCCTACTGCCAT	0.15	101	0.999	126	This study

^a Only primers with amplification efficiency between 90% and 110% were considered for RT-qPCR assays.

Supplementary Table 26. List of primers used for gene knockout experiments in *X. citri*.

Construction	Primer sequence (5'→3')
ΔXAC1768-69-F1	GGATCCCGCTATGGCAAATTCGGGTACTCC
ΔXAC1768-69-F2	CATATGTTGCGGGCCAGCCTGTAGAAC
ΔXAC1768-69-R1	CATATGCATCTGAATCCCTCCCAGGACG
ΔXAC1768-69-R2	AAGCTTATCGCCAGTGGCTGGTTGTGGAAT
ΔXAC1768-F1	GGATCCCGCTATGGCAAATTCGGGTACTCC
ΔXAC1768-F2	CATATGGCGCATGAACTGGTAAGTGC
ΔXAC1768-R1	CATATGCATCTGAATCCCTCCCAGGACG
ΔXAC1768-R2	AAGCTTAGATCGGTCTGGGTGTACCA
ΔXAC1769-F1	GGATCCAGGCCAATCAGTACGACACC
ΔXAC1769-F2	CATATGTTGCGGGCCAGCCTGTAGAAC
ΔXAC1769-R1	CATATGGATCGAACGCTGACCCACAG
ΔXAC1769-R2	AAGCTTATCGCCAGTGGCTGGTTGTGGAAT
ΔXAC1770-F1	GGATCCCTGGAATTTGCGCCCAACAACATCAAC
ΔXAC1770-F2	CATATGTATGGCGATCCACGATGAGCGC
ΔXAC1770-R1	CATATGCATTCTCCGGCTCCCCCTCG
ΔXAC1770-R2	AAGCTTCTTCCACGCGGACCGCAATG
ΔXAC1777-F1	GGATCCATACCGCGTTGAACCGTGCC
ΔXAC1777-F2	CATATGGAGCAGATGGAGGGCTGATGG
ΔXAC1777-R1	CATATGCATGGGTTACCTGGGAACATGC
ΔXAC1777-R2	AAGCTTGACCTGATACACCTTGGGCATTGAG

Supplementary Table 27. List of primers used for gene knockout verification by PCR and DNA sequencing.

Name	Primer sequence (5'→3')
1768-69_seqF	GGAATAGCAGCCTGTGAAATTTTCA
1768-69_seqR	TGTGCCACTGATACGGGTCTG
1768_seqF	GGAATAGCAGCCTGTGAAATTTTCA
1768_seqR	GTACTTCTAGTTTTTAGCGGGC
1769_seqF	ACGTTCAAATCAGTGCCATGT
1769_seqR	TGTGCCACTGATACGGGTCTG
1770_seqF	AACATCAATGATGTGTGGTCGATTT
1770_seqR	GCTTGCCGTCGAAGCTGA
1777_seqF	AGCGTTACGCCAGCTTCGACA
1777_seqR	TAAGTGCTGGATTGGCCTTCCCTT

References

1. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
2. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
3. Hsieh, Y. S. Y. & Harris, P. J. Xyloglucans of monocotyledons have diverse structures. *Mol. Plant* **2**, 943–965 (2009).
4. Hilz, H., de Jong, L. E., Kabel, M. A., Schols, H. A. & Voragen, A. G. J. A comparison of liquid chromatography, capillary electrophoresis, and mass spectrometry methods to determine xyloglucan structures in black currants. *J. Chromatogr. A* **1133**, 275–286 (2006).
5. Lerouxel, O. *et al.* Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting. *Plant Physiol.* **130**, 1754–1763 (2002).
6. Domon, B. & Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**, 397–409 (1988).
7. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
8. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
9. Glaser, F. *et al.* ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163–164 (2003).
10. Bitto, E. *et al.* The structure at 1.6 Å resolution of the protein product of the At4g34215 gene from *Arabidopsis thaliana*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1655–1661 (2005).
11. Till, M. *et al.* Structure and function of an acetyl xylan esterase (Est2A) from the rumen bacterium *Butyrivibrio proteoclasticus*. *Proteins Struct. Funct. Bioinforma.* **81**, 911–917 (2013).
12. Michalak, L. *et al.* A pair of esterases from a commensal gut bacterium remove acetylations from all positions on complex β-mannans. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7122–7130 (2020).
13. Correia, M. A. S. *et al.* Crystal structure of a cellulosomal family 3 carbohydrate esterase from *Clostridium thermocellum* provides insights into the mechanism of substrate recognition. *J. Mol. Biol.* **379**, 64–72 (2008).
14. Mølgaard, A., Kauppinen, S. & Larsen, S. Rhamnogalacturonan acetyl esterase elucidates the structure and function of a new family of hydrolases. *Structure* **8**, 373–383 (2000).
15. Henze, M. *et al.* Rational design of a glycosynthase by the crystal structure of β-galactosidase from *Bacillus circulans* (BgaC) and its use for the synthesis of N-acetyl lactosamine type 1 glycan structures. *J. Biotechnol.* **191**, 78–85 (2014).
16. Ohto, U. *et al.* Crystal structure of human β-galactosidase: Structural basis of G M1 gangliosidosis and morquio B diseases. *J. Biol. Chem.* **287**, 1801–1812 (2012).
17. Cartmell, A. *et al.* A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat. Microbiol.* **3**, 1314–1326 (2018).
18. Maksimainen, M. M., Lampio, A., Mertanen, M., Turunen, O. & Rouvinen, J. The crystal structure of acidic β-galactosidase from *Aspergillus oryzae*. *Int. J. Biol. Macromol.* **60**, 109–

- 115 (2013).
19. Wengelnik, K., Marie, C., Russel, M. & Bonas, U. Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction. *J. Bacteriol.* **178**, 1061–1069 (1996).
 20. Serrania, J., Vorhölter, F. J., Niehaus, K., Pühler, A. & Becker, A. Identification of *Xanthomonas campestris* pv. *campestris* galactose utilization genes from transcriptome data. *J. Biotechnol.* **135**, 309–317 (2008).
 21. Vorhölter, F. J. *et al.* The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *J. Biotechnol.* **134**, 33–45 (2008).
 22. Schatschneider, S. *et al.* Metabolic flux pattern of glucose utilization by *Xanthomonas campestris* pv. *campestris*: Prevalent role of the Entner-Doudoroff pathway and minor fluxes through the pentose phosphate pathway and glycolysis. *Mol. Biosyst.* **10**, 2663–2676 (2014).
 23. Schatschneider, S. *et al.* Establishment, in silico analysis, and experimental verification of a large-scale metabolic network of the xanthan producing *Xanthomonas campestris* pv. *campestris* strain B100. *J. Biotechnol.* **167**, 123–134 (2013).
 24. Yew, W. S. *et al.* Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from *Xanthomonas campestris*. *Biochemistry* **45**, 14582–14597 (2006).
 25. Seydel, A., Gounon, P. & Pugsley, A. P. Testing the ‘+2 rule’ for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol. Microbiol.* **34**, 810–821 (1999).
 26. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
 27. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
 28. Yu, C. S., Chen, Y. C., Lu, C. H. & Hwang, J. K. Prediction of protein subcellular localization. *Proteins Struct. Funct. Genet.* **64**, 643–651 (2006).
 29. Ferreira, R. M. *et al.* Unravelling potential virulence factor candidates in *Xanthomonas citri* subsp. *citri* by secretome analysis. *PeerJ* **4**, e1734 (2016).
 30. Pellock, S. J. *et al.* Three structurally and functionally distinct -glucuronidases from the human gut microbe *Bacteroides uniformis*. *J. Biol. Chem.* **293**, 18559–18573 (2018).
 31. Míguez Amil, S. *et al.* The cryo-EM Structure of *Thermotoga maritima* β -Galactosidase: Quaternary Structure Guides Protein Engineering. *ACS Chem. Biol.* **15**, 179–188 (2020).
 32. Ishikawa, K. *et al.* Crystal structure of β -galactosidase from *Bacillus circulans* ATCC 31382 (BgaD) and the construction of the thermophilic mutants. *FEBS J.* **282**, 2540–2552 (2015).
 33. Kong, C. *et al.* Complete genome sequence of strain WHRI 3811 race 1 of *Xanthomonas campestris* pv. *campestris*, the causal agent of black rot of cruciferous vegetables. *Mol. Plant-Microbe Interact.* **32**, 1571–1573 (2019).
 34. Lema, M., Cartea, M. E., Sotelo, T., Velasco, P. & Soengas, P. Discrimination of *Xanthomonas campestris* pv. *campestris* races among strains from northwestern Spain by *Brassica* spp. genotypes and rep-PCR. *Eur. J. Plant Pathol.* **133**, 159–169 (2012).
 35. Bogdanove, A. J. *et al.* Two new complete genome sequences offer insight into host and

- tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* **193**, 5450–5464 (2011).
36. Fargier, E. & Manceau, C. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathol.* **56**, 805–818 (2007).
 37. Richard, D. *et al.* Complete genome sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of solaneous plants, belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. *Genome Announc.* **5**, (2017).
 38. Dhakal, U., Dobhal, S., Alvarez, A. M. & Arif, M. Phylogenetic analyses of xanthomonads causing bacterial leaf spot of tomato and pepper: *Xanthomonas euvesicatoria* revealed homologous populations despite distant geographical distribution. *Microorganisms* **7**, 462 (2019).
 39. Jacques, M. A. *et al.* Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**, 163–187 (2016).
 40. Zhang, X. & Babadoost, M. Characteristics of *Xanthomonas cucurbitae* isolates from pumpkins and survival of the bacterium in pumpkin seeds. *Plant Dis.* **102**, 1779–1784 (2018).
 41. Ravanlou, A. & Babadoost, M. Development of bacterial spot, incited by *Xanthomonas cucurbitae*, in pumpkin fields. *HortScience* **50**, 714–720 (2015).
 42. Hayward, A. C. The hosts of *Xanthomonas*. in *Xanthomonas* Ch. 1 1–119 (Springer, Dordrecht, 1993). doi:10.1007/978-94-011-1526-1_1
 43. Rani, M., Weadge, J. T. & Jabaji, S. Isolation and characterization of biosurfactant-producing bacteria from oil well batteries with antimicrobial activities against food-borne and plant pathogens. *Front. Microbiol.* **11**, (2020).
 44. Garita-Cambronero, J., Palacio-Bielsa, A. & Cubero, J. *Xanthomonas arboricola* pv. *pruni*, causal agent of bacterial spot of stone fruits and almond: its genomic and phenotypic characteristics in the *X. arboricola* species context. *Mol. Plant Pathol.* **19**, 2053–2065 (2018).
 45. Cesbron, S. *et al.* Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* **6**, (2015).
 46. Fernandes, C., Blom, J., Pothier, J. F. & Tavares, F. High-quality draft genome sequence of *Xanthomonas* sp. strain CPBF 424, a walnut-pathogenic strain with atypical features. *Microbiol. Resour. Announc.* **7**, (2018).
 47. Wang, H., McTavish, C. & Turechek, W. W. Colonization and movement of *Xanthomonas fragariae* in strawberry tissues. *Phytopathology* **108**, 681–690 (2018).
 48. Kastelein, P. *et al.* Development of *Xanthomonas fragariae* populations and disease progression in strawberry plants after spray-inoculation of leaves. *Plant Pathol.* **63**, 255–263 (2014).
 49. Samanta, J. N., Mandal, K. & Maiti, S. A novel pathovar of *Xanthomonas axonopodis* causes gumming of Guggal (*Commiphora wightii*). *Eur. J. Plant Pathol.* **135**, 115–125 (2013).
 50. Jalan, N. *et al.* Comparative genomic analysis of *Xanthomonas axonopodis* pv. *citrumelo* F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. *J. Bacteriol.* **193**, 6342–6357 (2011).

51. Gottwald, T. R., Graham, J. H., Civerolo, E. L., Barrett, H. C. & Hearn, C. J. Differential host range reaction of citrus and citrus relatives to citrus canker and citrus bacterial spot determined by leaf mesophyll susceptibility. *Plant Dis.* **77**, 1004–1009 (1993).
52. Thieme, F. *et al.* Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* **187**, 7254–7266 (2005).
53. Abrahamian, P. *et al.* Molecular epidemiology of *Xanthomonas perforans* outbreaks in tomato plants from transplant to field as determined by singlenucleotide polymorphism analysis. *Appl. Environ. Microbiol.* **85**, (2019).
54. Ruh, M., Briand, M., Bonneau, S., Jacques, M. A. & Chen, N. W. G. *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *BMC Genomics* **18**, 670 (2017).
55. Darrasse, A. *et al.* 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 (2013).
56. EFSA Panel on Plant Health (PLH). Scientific Opinion on the pest categorisation of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans*. *EFSA J.* **12**, 3856 (2014).
57. Ruh, M., Briand, M., Bonneau, S., Jacques, M. A. & Chen, N. W. G. First complete genome sequences of *Xanthomonas citri* pv. *vignicola* Strains CFBP7111, CFBP7112, and CFBP7113 obtained using long-read technology. *Genome Announc.* **5**, (2017).
58. Moretti, C., Mondjana, A. M., Zazzerini, A. & Buonauro, R. Occurrence of leaf spot on cowpea (*Vigna unguiculata*) caused by *Xanthomonas axonopodis* pv. *vignicola* in Mozambique. *Plant Pathol.* **56**, 347 (2007).
59. Fonseca, N. P. *et al.* Analyses of seven new genomes of *Xanthomonas citri* pv. *aurantifolii* strains, causative agents of citrus canker B and C, show a reduced repertoire of pathogenicity-related genes. *Front. Microbiol.* **10**, (2019).
60. Jalan, N. *et al.* Complete genome sequence of *Xanthomonas citri* subsp. *citri* strain AW12879, a restricted-host-range citrus canker-causing bacterium. *Genome Announc.* **1**, (2013).
61. Midha, S. & Patil, P. B. Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. *Appl. Environ. Microbiol.* **80**, 6266–6279 (2014).
62. Brunings, A. M. & Gabriel, D. W. *Xanthomonas citri*: Breaking the surface. *Mol. Plant Pathol.* **4**, 141–157 (2003).
63. Da Silva, A. C. R. *et al.* Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463 (2002).
64. Al-Mousawi, A. H. Ultrastructural studies of a compatible interaction between *Xanthomonas campestris* pv. *malvacearum* and Cotton. *Phytopathology* **72**, 1222 (1982).
65. Phillips, A. Z. *et al.* Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLoS Genet.* **13**, (2017).
66. Sharma, V., Midha, S., Ranjan, M., Pinnaka, A. K. & Patil, P. B. Genome sequence of *Xanthomonas axonopodis* pv. *punicae* strain LMG 859. *J. Bacteriol.* **194**, 2395–2395 (2012).
67. Sharma, J. *et al.* Pomegranate bacterial blight: symptomatology and rapid inoculation

- technique for *Xanthomonas axonopodis* pv. *punicae*. *J. Plant Pathol.* **99**, 109–119 (2017).
68. Sutruedee, P. *et al.* Hemin transported protein of *Xanthomonas axonopodis* pv. *glycines* functions on leaf colonization and virulence on soybean. *African J. Microbiol. Res.* **7**, 4990–5003 (2013).
 69. G, K., Julian, S., David, S., Jerome, K. & E, K. Comparative pathogenicity studies of the *Xanthomonas vasicola* species on maize, sugarcane and banana. *African J. Plant Sci.* **9**, 385–400 (2015).
 70. Studholme, D. J. *et al.* Transfer of *Xanthomonas campestris* pv. *arecae* and *X. campestris* pv. *musacearum* to *X. vasicola* (Vauterin) as *X. vasicola* pv. *arecae* comb. nov. and *X. vasicola* pv. *musacearum* comb. nov. and desc. *Phytopathology* **110**, 1153–1160 (2020).
 71. Lang, J. M. *et al.* Detection and characterization of *Xanthomonas vasicola* pv. *vasculorum* (Cobb 1894) comb. nov. causing bacterial leaf streak of corn in the United States. *Phytopathology* **107**, 1312–1321 (2017).
 72. Nakato, G. V., Wicker, E., Coutinho, T. A., Mahuku, G. & Studholme, D. J. A highly specific tool for identification of *Xanthomonas vasicola* pv. *musacearum* based on five Xvm-specific coding sequences. *Heliyon* **4**, e01080 (2018).
 73. Wonni, I. *et al.* Analysis of *Xanthomonas oryzae* pv. *oryzicola* population in Mali and Burkina Faso reveals a high level of genetic and pathogenic diversity. *Phytopathology* **104**, 520–531 (2014).
 74. Lang, J. M. *et al.* A pathovar of *Xanthomonas oryzae* infecting wild grasses provides insight into the evolution of pathogenicity in rice agroecosystems. *Front. Plant Sci.* **10**, (2019).
 75. Niño-Liu, D. O., Ronald, P. C. & Bogdanove, A. J. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Mol. Plant Pathol.* **7**, 303–324 (2006).
 76. Ryan, R. P. *et al.* Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **9**, 344–355 (2011).
 77. Jaenicke, S. *et al.* Complete genome sequence of the barley pathogen *Xanthomonas translucens* pv. *translucens* DSM 18974T (ATCC 19319T). *Genome Announc.* **4**, (2016).
 78. Sapkota, S., Mergoum, M. & Liu, Z. The translucens group of *Xanthomonas translucens*: complicated and important pathogens causing bacterial leaf streak on cereals. *Mol. Plant Pathol.* **21**, 291–302 (2020).
 79. Peng, Z. *et al.* *Xanthomonas translucens* commandeers the host rate-limiting step in ABA biosynthesis for disease susceptibility. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20938–20946 (2019).
 80. Langlois, P. A. *et al.* Characterization of the *Xanthomonas translucens* complex using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic LAMP assays. *Phytopathology* **107**, 519–527 (2017).
 81. Cohen, S. P. *et al.* High-quality genome resource of *Xanthomonas hyacinthi* generated via long-read sequencing. *Plant Dis.* **104**, 1011–1012 (2020).
 82. Bansal, K. *et al.* Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated *Xanthomonas*. *bioRxiv* 453373 (2018). doi:10.1101/453373
 83. Fang, Y. *et al.* Genome sequence of *Xanthomonas sacchari* R1, a biocontrol bacterium isolated from the rice seed. *J. Biotechnol.* **206**, 77–78 (2015).
 84. Pieretti, I. *et al.* Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics* **13**, 658 (2012).

85. Agirre, J. *et al.* Privateer: Software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).
86. Cremer, D. & Pople, J. A. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1354–1358 (1975).
87. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
88. Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F. & Polikarpov, I. SAXSMoW 2.0: online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Sci.* **28**, 454–463 (2019).
89. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
90. Krissinel, E. & Henrick, K. ‘Protein interfaces, surfaces and assemblies’ service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html), inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
91. Cheng, W. *et al.* Structural insights into the substrate specificity of *Streptococcus pneumoniae* β (1,3)-galactosidase BgaC. *J. Biol. Chem.* **287**, 22910–22918 (2012).
92. Schröder, S. P. *et al.* Towards broad spectrum activity-based glycosidase probes: Synthesis and evaluation of deoxygenated cyclophellitol aziridines. *Chem. Commun.* **53**, 12528–12531 (2017).
93. Rojas, A. L. *et al.* Crystal structures of β -galactosidase from *Penicillium sp.* and its complex with galactose. *J. Mol. Biol.* **343**, 1281–1292 (2004).
94. Maksimainen, M. *et al.* Crystal structures of *Trichoderma reesei* β -galactosidase reveal conformational changes in the active site. *J. Struct. Biol.* **174**, 156–163 (2011).
95. de Melo, R. R. *et al.* Identification of a cold-adapted and metal-stimulated β -1,4-glucanase with potential use in the extraction of bioactive compounds from plants. *Int. J. Biol. Macromol.* **166**, 190–199 (2021).
96. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
97. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
98. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
99. Andrade, M. O., Farah, C. S. & Wang, N. The post-transcriptional regulator *rsmA/csrA* activates T3SS by stabilizing the 5' UTR of *hrpG*, the master regulator of *hrp/hrc* genes, in *Xanthomonas*. *PLoS Pathog.* **10**, e1003945 (2014).

Chapter 4

Type of chapter: Research Article

Current status: In preparation

Structural and functional characterization of a GH113 enzyme from a phytopathogenic Proteobacteria reveal molecular adaptations for a role in EPS cleavage

Isabela Mendes Bonfim^{1,2}, Marcelle Pandeló Martins¹, Mariane Noronha Domingues¹, Renan Augusto Siqueira Pirolla^{1,3,4}, Fabio Cesar Gozzo⁴, Priscila Oliveira de Giuseppe^{1,*} and Mário Tyago Murakami^{1,*}

¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPem), Zip Code 13083-970, Campinas, Sao Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

³ Current address: Waters Technologies Brasil, Barueri, São Paulo, Brazil

⁴ Institute of Chemistry, University of Campinas, Campinas, São Paulo, Brazil

*Correspondence to: priscila.giuseppe@lnbr.cnpem.br and mario.murakami@lnbr.cnpem.br

Abstract

The endo-mannanase activity has been widely studied in the glycoside hydrolase families GH5 and GH26. However, there are only a few reports about other families that also display this enzymatic activity, such as GH113 and GH134. Among the GH113 members, only three have been structurally characterized so far, all belonging to the Firmicutes phylum and currently displaying activity over plant cell walls heteromannans. In this work, we present the first crystal structure and kinetic characterization of a Proteobacterial GH113 member, the XagD enzyme from the phytopathogenic bacteria *Xanthomonas citri* pv. *citri* 306. This hypothetical protein is encoded by the last gene from the *xagABCD* operon, responsible for the biosynthesis of an exopolysaccharide (EPS) of unknown structure. Our results reveal that XagD can cleave heteromannans and manno-oligosaccharides, showing activity in substrates longer than mannotriose and tolerance for at least up to 28% of galactosyl decorations on the mannan backbone. Its endo-mannanase activity, low activity on plant-derived substrates and genomic context support the hypothesis that XagD has undergone molecular adaptations for using the Xag-EPS as substrate, hinting that the backbone of this not-yet characterized EPS might contain mannosyl β -1,4 linkages. In summary, this study shows how the intense remodeling of the substrate-binding cleft has allowed the emergence of biological roles as diverse as plant heteromannans degradation in GH113 enzymes from Firmicutes and EPS cleavage in those from Proteobacteria phytopathogens, expanding the current knowledge about the molecular evolution of the GH113 family.

Keywords: GH113, X-ray crystallography, glycoside hydrolase, exopolysaccharide, endo-mannanase

Introduction

The family GH113 was founded about 14 years ago with the characterization of the endo- β -1,4-mannanase AaManA from the thermo-acidophilic bacteria *Alicyclobacillus acidocaldarius* Tc-12-31¹. This family belongs to the clan GH-A and displays a classical $(\alpha/\beta)_8$ TIM-barrel fold and a retaining double-displacement catalytic mechanism^{1,2}. The biochemical activities already reported for its members include β -mannanase (EC 3.2.1.78), β -mannosidase (EC 3.2.1.25) and reducing-end exo- β -mannosidase (EC 3.2.1.-). Depending on their specificity, GH113 enzymes can recognize and cleave homo- or heteromannans, as well as mannooligosaccharides, generating or not transglycosylation products^{1,3-8}.

Currently, this family comprises 1581 members, from each only 20 have been functionally characterized according to the Carbohydrate Active Enzymes (CAZy) database⁹. Recently, a phylogenetic analysis of this family distributed its members into five clades, mostly associated with the taxonomy of their source species³. Notably, all the GH113 enzymes structurally characterized so far belong to clade 1^{1,3,6,10}, which mainly comprises representatives from the Firmicutes phylum, reinforcing the importance of the structural characterization of members from other clades to a better comprehension of the molecular evolution of this family.

The clade 2 contains members from the Proteobacteria phylum³, some of which are found in phytopathogenic bacteria from the *Xanthomonas* spp. genus. Thus, we chose a clade 2 member encoded by *Xanthomonas citri* pv. *citri* strain 306 (hereafter *X. citri* 306), as a model of study to investigate the structure and function of a proteobacterial GH113 enzyme, seeking for a better comprehension of the molecular adaptations occurred before the split of the Proteobacteria and Firmicutes phyla. The target protein is named XagD¹¹ and its gene (XAC3522) is the only found in the *X. citri* 306 genome to encode a GH113 family member. Interestingly, the *xagD* gene composes an operon with the genes *xagABC*, responsible for the biosynthesis of a not-yet characterized exopolysaccharide (EPS) that contributes to biofilm formation¹¹. In *Xanthomonas campestris*, deletion of the genes *xagABC*, which encodes for glycosyl transferases, considerably affected the biofilm formation and maintained the cells in the planktonic state¹¹. In the same study, however, the deletion of the gene *xagD* did not impaired the biofilm formation, keeping elusive the biological role of this protein. Recently, this protein was assigned as an endo- β -1,4-mannanase³, suggesting it can display a role in EPS cleavage during biofilm formation.

This work presents the biochemical and structural characterization of the enzyme XagD from *X. citri* 306, providing insights about substrate preference and structural adaptations that likely shaped this enzyme to recognize and cleave a putative β -1,4-mannose-based EPS that is critical for biofilm formation in Proteobacteria phytopathogens from the *Xanthomonas* genus, contrasting with the biological roles previously proposed for the GH113 enzymes characterized in Firmicutes, mainly associated to the depolymerization of plant cell wall heteromannans.

Results and Discussion

XagD shows activity in polymeric heteromannans

To investigate the biochemical activity of XagD (XAC3522, GenBank AAM38365.1), we performed a substrate screening with 9 synthetic and 23 polymeric carbohydrates (Supplementary Tables 1, 2). XagD showed no activity on the pNP-derived substrates, indicating that this enzyme does not act as β -mannosidase, but it was active on the polymeric substrates galactomannan and glucomannan, with a residual activity on the insoluble mannan from ivory nut (Supplementary Table 2).

XagD has an optimum temperature of 40 °C and optimum pH between 5.5 and 6.5 (Supplementary Figure 1), with the highest specific activity on guar galactomannan 21%, about 2.6 higher compared to the other heteromannans tested (Figure 1A). Regarding the polymeric substrates evaluated, it was observed that the products formed by XagD consist mostly of mannobiose (M2) and mannotriose (M3) in 22 h reactions (Figure 1B-F). For the substrate guar galactomannan containing 38% galactose substitutions, no activity was detected, indicating that increasing the frequency of galactosyl substitutions reduces the recognition sites of the enzyme on the substrate. The decrease in activity in galactomannans with high content of galactosyl decorations (38%) have been previously observed in other GH113 enzymes and is likely a conserved feature of this family (Supplementary Table 3).

Regarding the cleavage pattern of manno-oligosaccharides, XagD cleaved a portion of the total mannopentaose present in the reaction, generating M2 and M3, but much less efficiently compared to the cleavage of mannohexaose, which was fully converted into M2, M3, and M4, indicating that XagD might have at least 6 subsites for substrate binding (Figure 2). XagD showed residual activity on mannotetraose and no activity on mannobiose, mannotriose, 6-1- α -D-galactosyl-mannotriose (G-M3) and 6-3, 6-4- α -D-galactosyl-mannopentaose (GG-M5), showing that its substrate-binding site is intolerant to α -D-galactosyl decorations simultaneously present in the 3 and 4 positions of M5 (Figure 2).

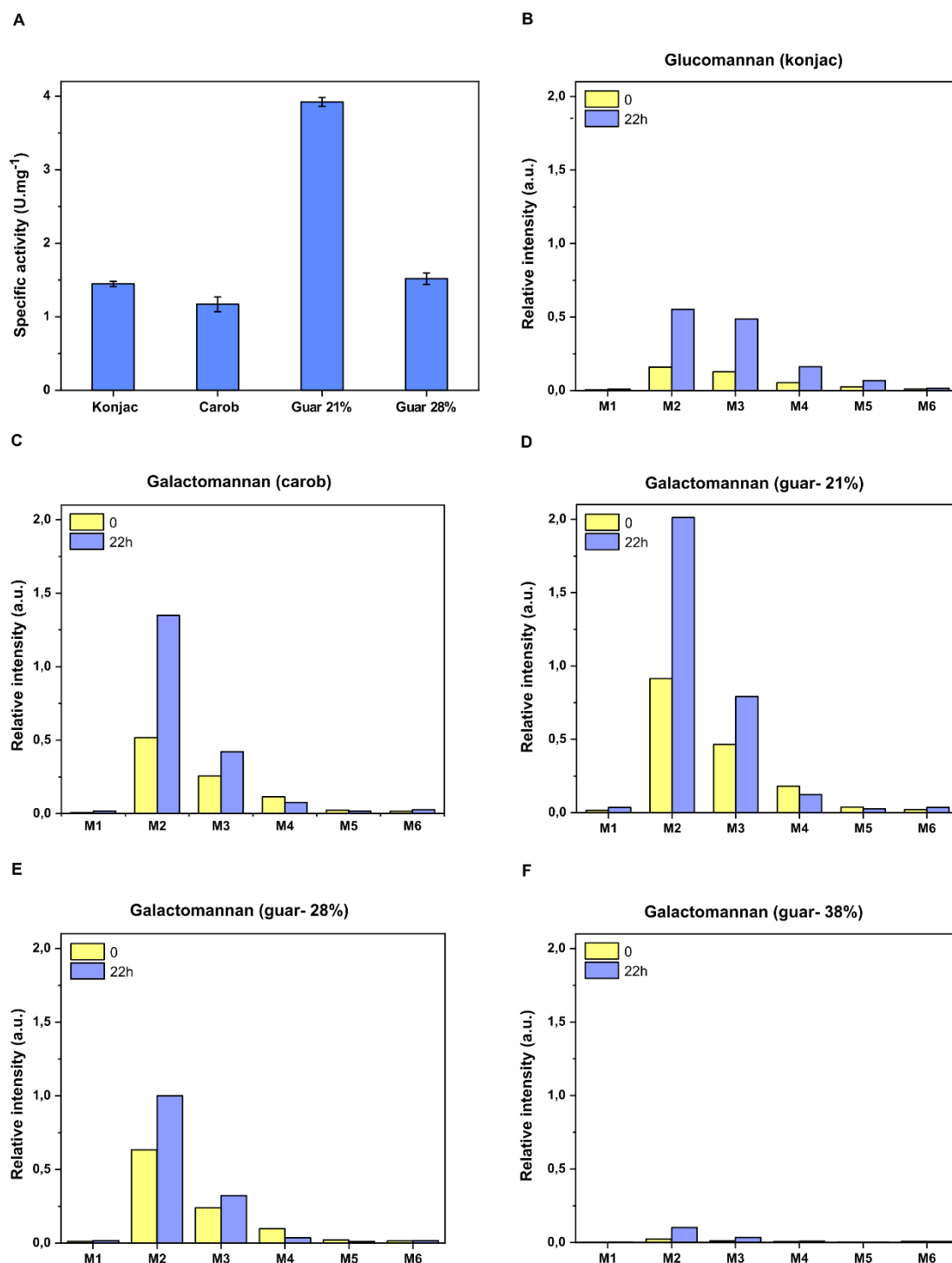


Figure 1. Specific activity and profile of the cleavage pattern of mannan polysaccharides by the endo-mannanase XagD. **A)** Specific activity of XagD on different mannan polysaccharides assessed by the 3,5-dinitrosalicylic acid method¹². **B-F)** Plots show the intensity values of the analytes (mannan-oligosaccharides) divided by the intensity of the internal standard (xylotriase) detected by mass spectrometry. The bars in yellow represent the control reactions (no enzyme), while the bars in blue represent the product of XagD reactions incubated with the substrates glucomannan (konjac), galactomannan (carob) and guar galactomannan (with 21, 28 or 38% galactose substitutions) for 22 hours.

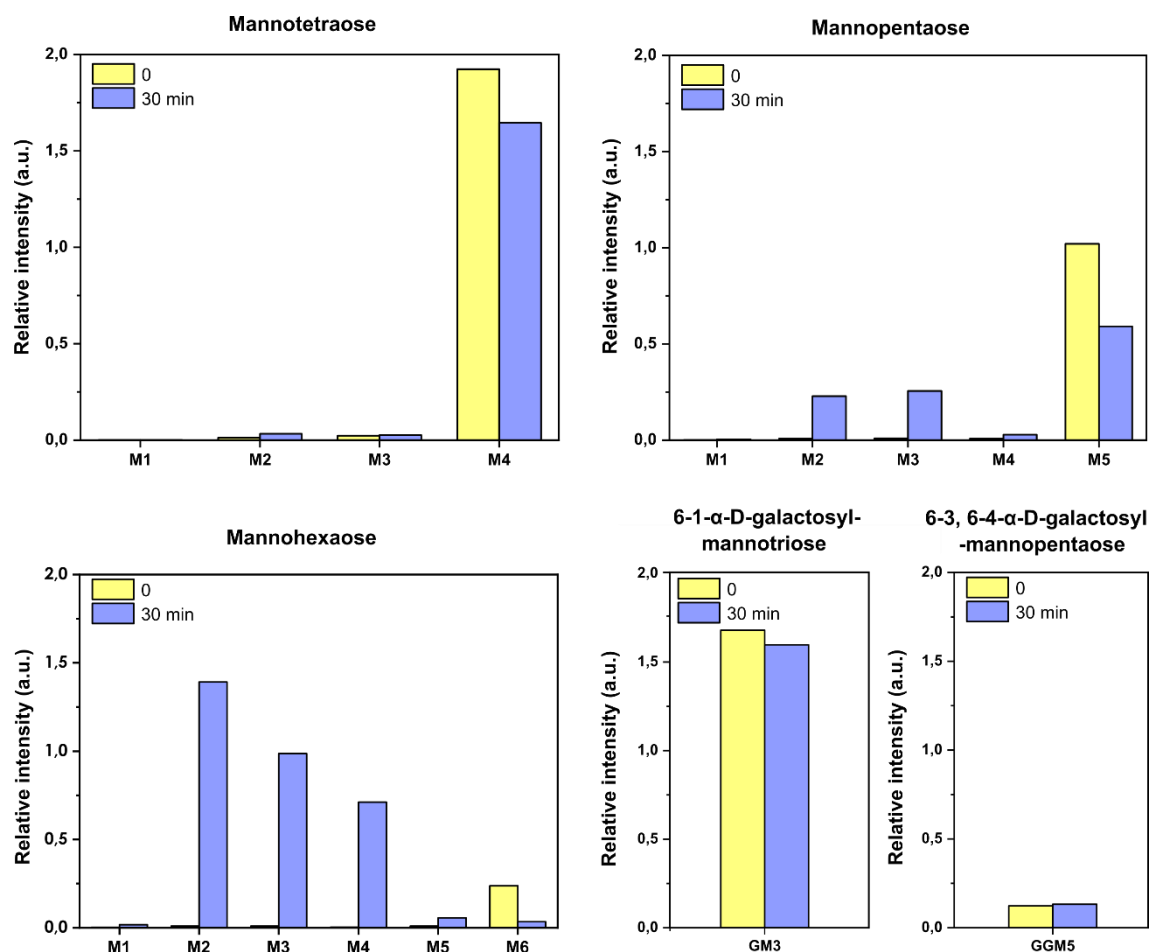


Figure 2. Profile of the cleavage pattern of manno-oligosaccharides by the endo-mannanase XagD.

The graphs show the intensity values of the analytes (manno-oligosaccharides) divided by the intensity of the internal standard (xylotriase). The bars in yellow represent the control reactions (no enzyme) and the bars in blue represent the product of the reactions of XagD incubated with the substrates mannopentaose (M5), mannohexaose (M6), 6-1- α -D-galactosyl-mannotriose (GM3) and 6-3, 6-4- α -D-galactosyl-mannopentaose (GG-M5) for 30 minutes.

In substrate saturation curve assays, the kinetic parameters of XagD on manno-oligosaccharides were estimated. Consistent with the profile observed in the cleavage pattern assays, XagD showed a threefold lower affinity for M5, with a K_m equal to 17.8 mM, compared to M6, with a K_m of 5.5 mM (Table 1, Supplementary Figure 2). The k_{cat} values obtained were also considerably low: 0.175 s⁻¹ for M5 and 0.418 s⁻¹ for M6, indicating that the enzyme catalyzes the cleavage of less than one molecule of these substrates per second. XagD displayed a k_{cat}/K_m constant 8-fold higher for M6 (0.08 s⁻¹mM⁻¹) compared to M5 (0.01 s⁻¹ mM⁻¹), indicating that it is more specific for M6 than for M5.

Table 1. Activity comparison between XagD, *Aa*ManA, *Ax*Man113A, *Ba*Man113A (WT and N236Y) and *Pc*Man113 (WT and against manno-oligosaccharides (M2-M6).

Protein	M2 k_{cat} (s ⁻¹)	M2 K_m (mM ⁻¹)	M2 (k_{cat}/K_m) (s ⁻¹ mM ⁻¹)	M3 k_{cat} (s ⁻¹)	M3 K_m (mM ⁻¹)	M3 (k_{cat}/K_m) (s ⁻¹ mM ⁻¹)	M4 k_{cat} (s ⁻¹)	M4 K_m (mM ⁻¹)	M4 (k_{cat}/K_m) (s ⁻¹ mM ⁻¹)	M5 k_{cat} (s ⁻¹)	M5 K_m (mM ⁻¹)	M5 (k_{cat}/K_m) (s ⁻¹ mM ⁻¹)	M6 k_{cat} (s ⁻¹)	M6 K_m (mM ⁻¹)	M6 (k_{cat}/K_m) (s ⁻¹ mM ⁻¹)	Reference
XagD	NC	NC	NC	NC	NC	NC	-	-	RA	0.175 ± 0.017	17.8 ± 2.07	0.01	0.418 ± 0.021	5.51 ± 0.35	0.08	This work
<i>Aa</i> ManA	NC	NC	NC	-	-	NQ	8.83	4.55 ± 0.58	1.94	8.53	1.15 ± 0.11	7.02	8.51	0.95 ± 0.11	7.99	1
Man113A	NC	NC	NC	-	-	6.54	-	-	29.70	-	-	72.70	-	-	75.70	8
<i>Ax</i> Man113A	0.098	74.3 ± 4.4	0.0013	0.853	29.1 ± 2.3	0.03	0.979	12.7 ± 1.1	0.08	1.065	5.3 ± 0.4	0.20	1.508	3.9 ± 0.3	0.39	6
<i>Ba</i> Man113A WT	1.41	23.53	0.06	7.70	2.83	2.72	4.82	1.66	2.9	5.14	1.41	3.66	-	-	NQ	10
<i>Ba</i> Man113A N236Y	NC	NC	NC	0.84	4.17	0.2	0.57	1.72	0.3	1.01	3.74	0.3	-	-	NT	10
<i>Pc</i> Man113 WT	2.36	66.1 ± 5.7	0.035	7.38	35.4 ± 4.1	0.2	26.48	5.1 ± 3.4	5.19	31.83	4.7 ± 0.5	6.77	26.88	8.2 ± 0.8	3.28	4
<i>Pc</i> Man113 N246Y	-	-	NT	-	-	NT	26.48	4.5±1.7	5.89	34.75	1.8 ± 0.1	19.30	25.62	4.3±0.4	5.96	4

NC: no cleavage detected; NQ: cleavage is detected, but not quantified; NT: activity not tested; RA: residual activity detected

Saturation curves were also performed for several mannan polysaccharides (Supplementary Figure 3), except for glucomannan (konjac) due to the low solubility of this substrate in the concentration range required to reach enzyme saturation. By comparing the kinetic parameters, we observed that XagD presents similar specificity (k_{cat}/K_M) over the galactomannans tested (Table 2). The higher k_{cat} values observed for the substrates galactomannan carob (24%) and galactomannan guar (28%) correlates with higher K_M values compared to galactomannan guar (21%), indicating that an increase in the frequency of galactosyl decorations from 21% to 24-28% along the mannan backbone is detrimental for substrate affinity, but favorable for enzyme turnover.

Table 2. Comparison of kinetic parameters of the XagD enzyme on the substrates galactomannan carob and galactomannan guar. The content of galactose substitutions is indicated in parentheses. K_M and k_{cat} are presented as mean \pm SD and were estimated from an average curve estimated from three independent experiments (n=3).

Substrate	K_M (mg. mL ⁻¹)	k_{cat} (s ⁻¹)	k_{cat}/K_M (mL s ⁻¹ mg ⁻¹)
Galactomannan guar (21%)	4.80 \pm 0.490	1.81 \pm 0.20	0.38
Galactomannan carob (24 %)	6.63 \pm 0.61	1.72 \pm 0.15	0.26
Galactomannan guar (28%)	6.16 \pm 1.77	1.32 \pm 0.48	0.21

XagD overall structure

To get insights about the structural features that dictate the substrate recognition mechanism and the biological role of XagD, we solved its crystallographic structure by single anomalous diffraction (SAD) using the iodine soaking method (data named XagD_iodine). The native structure was also determined (named XagD_native) (Supplementary Table 4). XagD crystals showed one monomer in the asymmetric unit and no probable quaternary structure was detected in the crystal lattice using PDBePISA¹³. Small-angle X-ray scattering (SAXS) analysis resulted in a radius of gyration (R_g) of 1.98 \pm 0.003 nm, similar to the crystallographic monomer, confirmed by its *ab initio* envelope (Supplementary Figure 4A-B) and in agreement with dynamic light scattering (DLS) assays (R_h = 2.68 \pm 0.48 nm), confirming the monomeric state of XagD in the evaluated conditions. The monomeric form can be associated with the shortening and structural modification of loop A133-P138 in XagD since it has an essential role in interactions that stabilize the dimeric structure of the GH113 enzyme BaMan113A¹⁰ (Supplementary Figure 4C-D).

XagD shows the classic $(\beta/\alpha)_8$ barrel, and the conserved residues of the clan GH-A were identified by the structural superposition of XagD with the most similar and unique GH113 crystallographic structures known so far, the β -1,4-mannanases AxMan113A from *Amphibacillus xylanus* (23% of sequence identity)⁶, BaMan113A from *Bacillus* sp. (22% of sequence identity)¹⁰, and AaManA from *Alicyclobacillus acidocaldarius* (23% of sequence identity)⁸, according to the Dali server¹⁴ (Figure 3A).

Although XagD active site conserves most of the residues expected in the clan GH-A, including the catalytic residues E257 and E169, the positions occupied by H122 and T168 in XagD are not fully conserved in the family GH113¹ (Figure 3A). The presence of H122 in enzymes from clade 2 seems to help in catalytic nucleophile E257 stabilization through ionic interactions, which is not observed in the other GH113 structures from clade 1 known so far (Figure 3A). Regarding T168, except for XagD, the other known GH113 structures retain a cysteine residue at this position (Figure 3A), which possibly affect the position of the adjacent nucleophile during hydrolysis, since its mutation for alanine reduced the activity of AaManA¹ and Man113A⁸. The substitution of the conserved cysteine for T168 in XagD is intriguing mainly due to the unexpected orientation of the threonine's side chain oxygen, which is not pointing to the same direction of the cysteine side chain from the homologous structures, according to the electron density map (Supplementary Figure 5).

The residues K120, S219, and Y221 are well described in clan GH-A for contributing to the correct positioning and ionization state of the two catalytic glutamic acids¹ (Figure 3A and B). Although the orientation of S219 in XagD_{native} is conserved, its orientation in XagD_{iodine} is modified. While the interaction with acid/base E169 is maintained, the interaction with the nucleophile E257 is prevented in XagD_{iodine} structure (Figure 3C). The unusual orientation of S219 in XagD_{iodine} might be influenced by the movement of loop L195-M201, induced by the binding of an iodine ion at its vicinity (Supplementary Figure 6A).

To identify the positive and negative subsites of XagD, the enzyme was co-crystallized with manno-oligosaccharides M2, M3 and M4. Unfortunately, no ligands were visualized in the active site, except for a CHES molecule from the crystallization condition found in all collected data, including the XagD_{iodine} dataset (Supplementary Figure 7). Then, the identification of subsites was realized by structural superposition with the GH113 crystallographic structures in complex with manno-oligosaccharides reported so far.

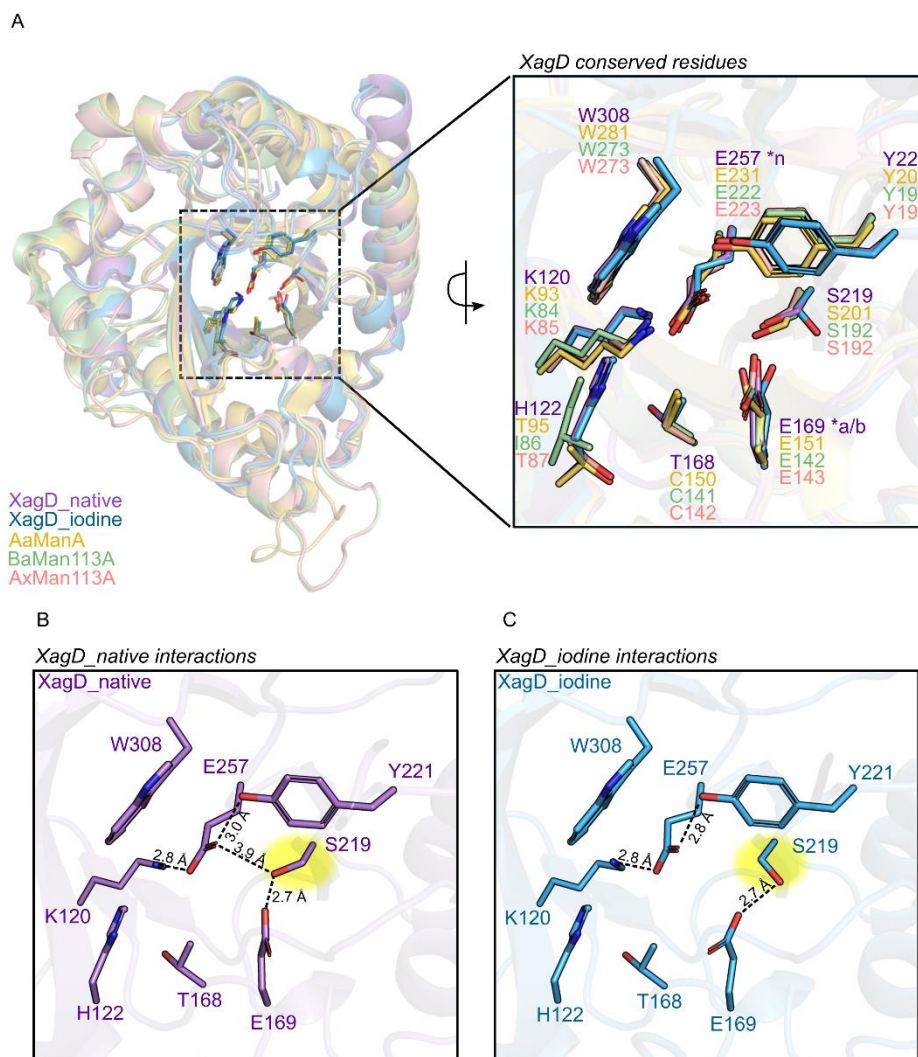


Figure 3. XagD conserved residues. A) Structure superimposition of XagD_iodine (blue), XagD_native (purple), AaManA (PDB code 4CD8, rmsd 1.12 Å, yellow), BaMan113A (PDB code 7DV7, rmsd 1.12 Å, green), and AxMan113A (PDB code 5YLH, rmsd 1.30 Å, pink). The zoom indicates active site conserved residues. “n” and “a/b” represent the nucleophile and acid/basic catalytic glutamic, respectively. B) Representation of catalytic residues (E169 and E257) interactions in the active site of XagD_native. The usual orientation of S219 is highlighted in yellow. C) Representation of catalytic residues (E169 and E257) interaction in the active site of XagD_iodine highlighting the unusual orientation of S219 and the loss of interaction with E257. Dashed lines represent interatomic distances.

Iodine ions induced the closure of a lid over the substrate-binding subsites +1 and -1 in XagD

XagD showed the deep substrate-binding cleft typically found in GH113 members^{1,6,10} (Figure 4A). By comparing the XagD_native with the XagD_iodine structures (Figure 4A and B), we observed a conformational change of the W129 side chain that resulted in the formation of a hydrophobic lid over the -1 and +1 subsites of XagD_iodine cleft (Figure 4C). This conformational change apparently was induced by the binding of iodine ions in the W129

vicinity (Supplementary Figure 6B), so it remains to be seen whether they have biological meaning, i.e., if they are induced, for example, by substrate binding, or whether they are simply a methodological artifact.

The lid formation is also observed and reported in the other GH113 structures known so far (Supplementary Table 5, Figure 4C) and might have implications on substrate recognition and substrate specificity since it can preclude the binding of decorated moieties in specific subsites. For example, in XagD lid closure occurs between -1 and +1 subsites whereas in AaManA it is over -2 and -1 subsites (Figure 4C). Interestingly, the residue assuming the role of the hydrophobic lid in XagD is in a different position from the other GH113 structures reported. In these cases, hydrophobic residues replacing XagD-Q276 in AaManA (Y247), BaMan113A (W238) and AxMan113A (W239) act as lids, assuming a closed conformation upon substrate binding (Figure 4C). Lid interactions with the other side of the catalytic cleft seems to stabilize the closed conformation in some structures, as supported by the conformational changes observed for XagD-Q276 and BaMan113A-W94 upon lid closure. In XagD, the lid closure involves not only a local change of W129 rotamer orientation but also a modification into the secondary structure of the W129-D137 loop (Supplementary Figure 6B). This modification correlates with the conformational change of the loop harboring XagD-Q276, which has a higher flexibility compared to other parts of the enzyme, not only in XagD but also in other GH113 structures, according to B-factor analysis (Supplementary Figure 8).

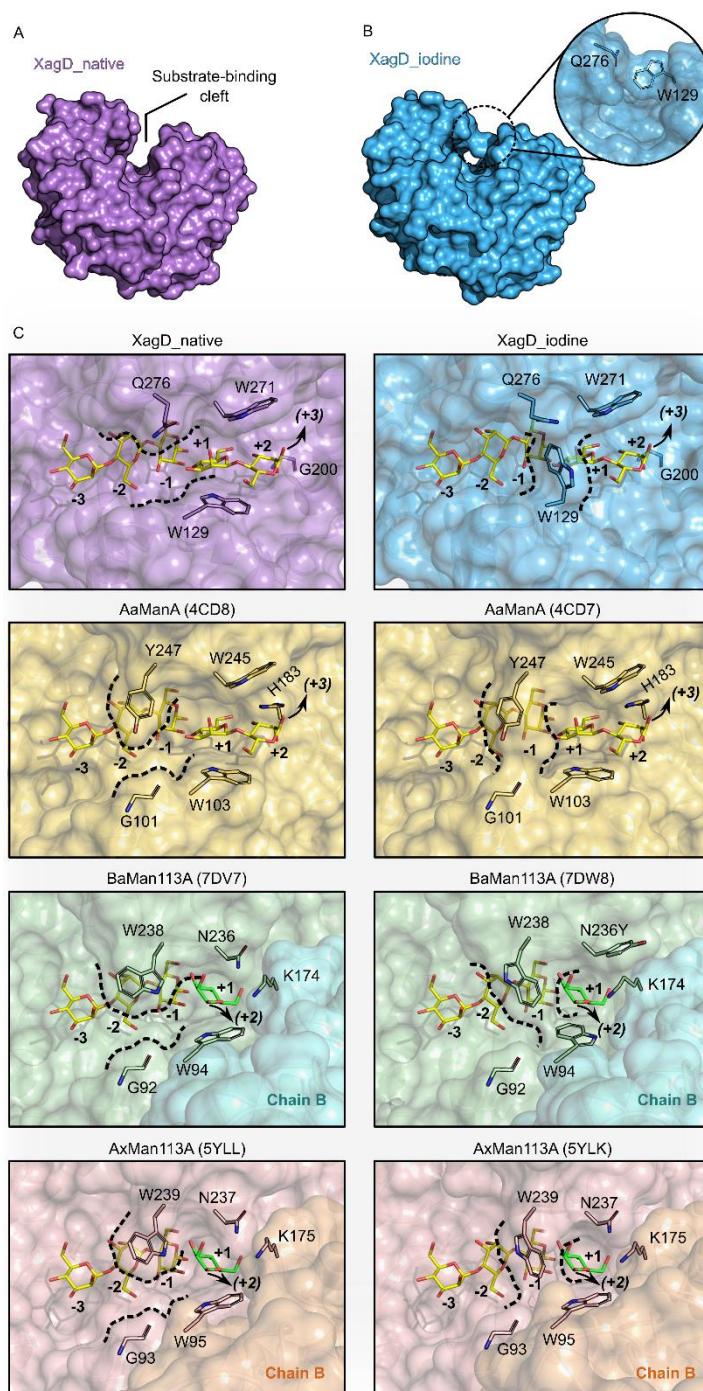


Figure 4. Lid forming mechanism in XagD diverges from other GH113 enzymes. **A)** Side view of substrate binding-cleft in XagD_{native}. **B)** Side view of lid formation above active site of XagD_{iodine}. **C)** Comparisons of open and closed conformations of the lid in XagD and GH113 structures deposited in PDB (respective PDB code is described in each image). In the monomeric enzymes XagD (purple and blue) and AaManA (yellow), the arrow indicates the location of +3 subsite. In the dimeric enzymes BaMan113A (chain A in light green and chain B in light blue) and AxMan113A (chain A in pink and chain B in orange), the region corresponding to the +2 subsite in XagD and AaManA is blocked by the chain B of the dimer, leading to an alternative pathway for +2 mannose accommodation, as indicated by the arrow. The delimitation of the lid is represented by the black dashed lines on each structure. Mannose, manno- and mannotriose coordinates were obtained from the crystallographic structures of AxMan113A (PDB code 5YLK, rmsd 1.20 Å), AaManA (PDB code 4CD7, rmsd 1.12 Å) and AxMan113A (PDB code 5YLL, rmsd 1.21 Å), respectively.

Modifications in specific residues decrease affinity to mannose moieties at positive and negative subsites in XagD

In the GH113 structures known so far, the +2 subsite is located in different regions depending on the enzyme's oligomeric state. For example, in *AxMan113A* and *BaMan113A*, the dimer interface blocks a region corresponding to the +2 subsite observed in monomeric enzymes such as *AaManA* (PDB code 4CD7) and XagD (Figure 4C). This blockage offers an alternative pathway for substrate binding, which involves a conformational change in the substrate, as indicated by the different positioning of mannose in the +1 subsite of the dimeric *AxMan113A* (PDB code 5YLK) compared to the monomeric *AaManA* (PDB code 4CD7) (Figure 4D, Supplementary Figure 9).

Besides this difference, the dimeric enzymes *BaMan113A* (N236 and K174)¹⁰ and *AxMan113A* (N237 and K175)⁶ conserve an asparagine and a lysine residue that contributes to the substrate affinity at +1 subsite, but are not conserved in *AaManA* and XagD (Supplementary Figure 9). The asparagine residue provides a hydrogen bond to the O6 atom from +1 mannose whereas the lysine residue makes an electrostatic interaction with E109 from the neighboring chain (*AxMan113A*), stabilizing the dimer interface and the proper positioning of the E109 side chain, which makes a hydrogen bond with the O6 atom of +1 mannose (Supplementary Figure 9A and B). Similarly, to *AaManA* (W245), the asparagine residue is replaced by a tryptophan in XagD (W271, Supplementary Figure 9C and D). This replacement decreases the number of polar interactions with the +1 mannosyl moiety while provides a hydrophobic platform for +2 mannosyl binding in a surface region that is blocked in the dimeric enzymes (Figure 4D, Supplementary Figure 9C and D). Interestingly, XagD and *AaManA* are not active on mannobiose, as well as the mutant *BaMan113A* N236Y¹⁰, indicating that the loss of polar interactions with the substrate at the +1 subsite, induced either by the change in the oligomeric state or by the asparagine to tryptophan/tyrosine replacement prevents the activity of GH113 enzymes over mannobiose, playing an important role in the determination of their substrate specificity.

Regarding the lysine residue (*AxMan113A*-K175), it is substituted by a histidine residue in *AaManA* (H183), providing more room at the +2 subsite and hydrogen bonds to both the O3 atom from the +2 mannose residue and to the O6 atom from +1 mannose moiety (Supplementary Figure 9C). In XagD, the lysine is replaced by a glycine (G200), which opens space for substrate binding at +2 subsite but without providing interactions for substrate

stabilization at the +1 and +2 subsites (Supplementary Figure 9D). Thus, compared to *AaManA*, XagD provide less hydrogen bonds to the substrate at the +1 and +2 subsites, especially due to the H183G replacement in XagD, which likely explain why XagD has only a residual activity over mannotetraose whereas *AaManA* recognizes and cleaves with efficiency this substrate.

Although XagD barely cleaves M4, it is progressively more active over M5 and M6 substrates, being eight-fold more specific to M6 compared to M5 and generating M3 as one of the main products of M6 cleavage (Figure 2, Table 1), which indicates the existence and the importance of an additional +3 subsite for substrate binding in this enzyme. The +3 subsite region is poorly conserved between XagD and *AaManA*, except for the residue W271 (XagD), which is also conserved in *AaManA* (W245) and seems to provide a platform that is sufficiently large to provide hydrophobic contacts for substrate binding at both the +2 and the +3 subsites (Figure 4D).

The negative subsites of XagD are marked by intense remodeling compared to the GH113 enzymes from clade 1. Some of the observed replacements likely result in loss of polar interactions with mannose residues, such as at the positions F324, and A130 of XagD, which correspond respectively to conserved tyrosine and arginine residues in the homologous enzymes (Figure 5). The substitution of aromatic residues by K49 and S273 result in loss of hydrophobic interactions at -2 subsite in XagD, likely decreasing substrate affinity at this subsite. On the other hand, the replacement of a conserved asparagine for W124 in XagD seems to provide a hydrophobic platform for substrate-binding at the -3 subsite while maintaining the hydrogen bond with the mannose residue in the -2 subsite. Interestingly no direct interactions of mannose residues with the -3 and -4 subsites are observed for *AxMan113A* (PDB code 5YLI)⁶, contrasting with this possible stacking interaction provided by W124 at XagD -3 subsite, which might compensate the loss of interactions at the -2 subsite (Figure 5).

Of note, the replacement of a conserved arginine (R96, R95, and R104 in *AxMan113A*, *BaMan113A* and *AaManA*, respectively) for an alanine (A130) in XagD might have contributed to the lower activity observed for XagD on manno-oligosaccharides in comparison with its characterized counterparts (Table 1). This arginine residue makes a hydrogen bond with the O2 and O3 atoms from the -1 mannosyl residue and a salt bridge with the catalytic acid/base glutamic. Its mutation to alanine causes a completely loss of activity in *AxMan113A*³ and *Man113A*⁸. Interestingly, this strategic residue is naturally substituted by an alanine in XagD (A130), causing the loss of the critical interactions mentioned above, likely being one of the

key replacements related to the lower activity found for XagD on manno-oligosaccharides compared with the other characterized GH113 members (Table 1).

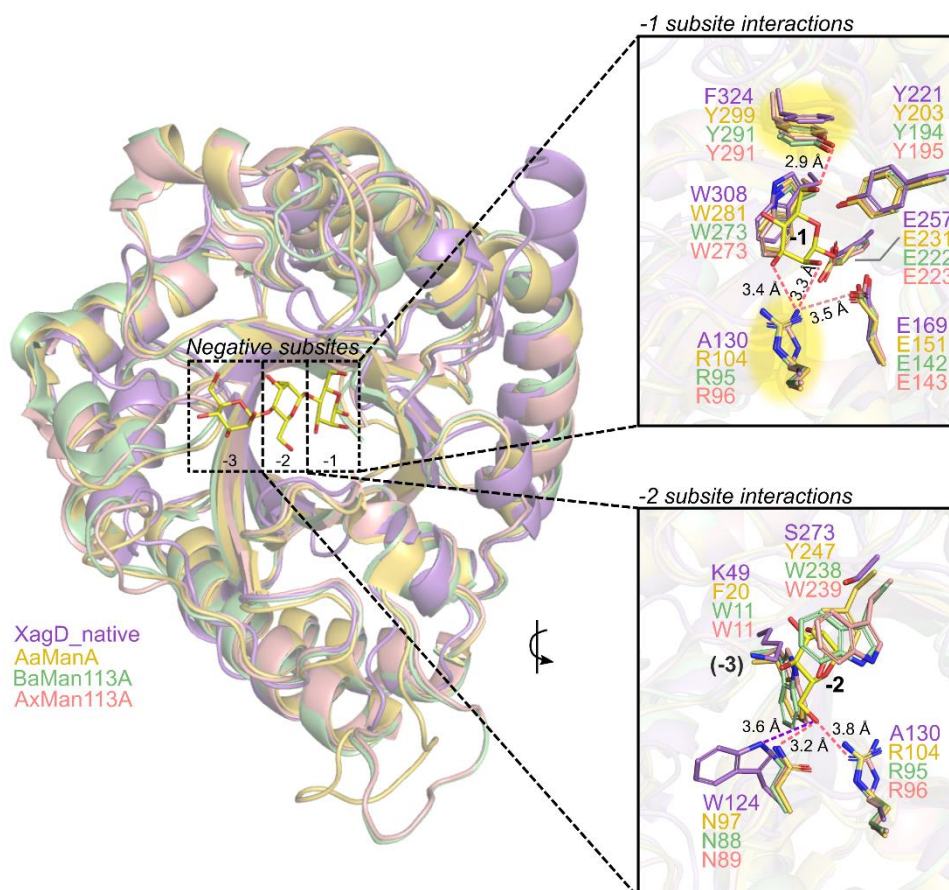


Figure 5. View of XagD negative subsites interactions. Cartoon representation of GH113 crystallographic structures superimposed, highlighting the negative subsites defined based on the structure of AxMan113A complex with M6 (PDB code 5YLL). The top and downright panels describe the residues that delineate the -1 and -2 subsites of XagD (purple), AaManA (PDB code 4CD8, r.m.s.d. 1.1 Å, yellow), BaMan113A (PDB code 7DV7, r.m.s.d. 1.2 Å, green) and AxMan113A (PDB code 5YLH, r.m.s.d. 1.2 Å, pink), respectively. The region corresponding to the -3 subsite is indicated in parenthesis. Non-conserved residues in subsite -1 of XagD are highlighted in yellow. Manno-oligosaccharides coordinates (sticks, yellow C atoms) are from the AxMan113A crystal structure (PDB code 5YLL, r.m.s.d. 1.2 Å). Dashed lines represent interatomic distances in AxMan113A (hydrogen bonds and salt bridge are colored in pink and light pink, respectively) and XagD_native (the interatomic distance compatible with a hydrogen bond is purple colored).

Remodeling of substrate-binding cleft in XagD correlates with a possible role in EPS cleavage

The replacements of conserved residues that are important for manno-oligosaccharides affinity in the negative and positive subsites indicate either a change in the substrate specificity of XagD or a strategy to decrease enzyme activity, since reaction velocity and substrate affinity are interdependent. These hypotheses are supported by the lower activity of XagD in mannans polysaccharides and oligosaccharides compared to others GH113 members biochemically characterized so far (Table 1 and Supplementary Table 3).

According to its genomic context, XagD can be involved in the cleavage of a EPS synthesized by family 2 glycosyl transferases (GT2) encoded by neighboring genes, as suggested for GH113 enzymes from clade 5, whose genes are also commonly associated to GT2 enzymes³. Thus, the modifications occurred in the substrate-binding cleft of XagD might have favored a biological role unrelated with plant mannan degradation but associated with the cleavage of a mannan-like bacterial EPS in the context of biofilm assembly. The biological role of GH enzymes present in EPS synthesis operons is still partially understood, but some proposed biological functions include EPS length control, generation of a secretion-competent form of the EPS and degradation of excess EPS in the periplasm^{15–17}. Therefore, future studies aiming to isolate and determine the structure of the Xag-EPS, as well as, to characterize the XagD interaction and activity on this possible substrate will be instrumental and complementary to our results for a better understanding of XagD biological role.

Conclusion

In this study we presented the first structural and functional characterization of a GH113 member from the Proteobacteria phylum, XagD, expanding the current knowledge about the molecular evolution of this family for beyond the Firmicutes phylum. Our work revealed that XagD is a monomeric enzyme with a substrate-binding cleft that suffered intense remodeling compared to its Firmicutes counterparts, resulting in the loss of several interactions with manno-oligosaccharides substrates, decreasing both affinity and turnover, which might have been an adaptation for the limited cleavage of a mannan-like EPS, whose structure is yet to be determined but whose role in biofilm formation has been proved to be essential in *Xanthomonas* spp.

Methods

Cloning, protein expression and purification

The XAC3522 nucleotide sequence (GenBank AAM38365.1) is predicted to be anchored in the outer membrane according to sequence-based prediction analysis (Supplementary Table 5). The gene was amplified from the genomic DNA of *X. citri* 306 strain using standard methods (primer forward - GAATTCCATATGGCCGCGCCACCTGGATG and reverse - CTCGAGTTAGCGCGCAAGTACCTGCTGCG). The signal peptide was removed, and the PCR-amplified gene fragments were cloned into pET28a vector after digestion with NdeI and XhoI restriction enzymes, with insertion of a 6xHis-tag aiming further purification steps.

Next, the pET28a-XagD plasmid was transformed into thermocompetent *Escherichia coli* strain BL21(DE3) cells and grown in Terrific Broth (TB) medium (1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.4% (w/v) glycerol, 0.017 M sodium phosphate monobasic and 0.072 M sodium phosphate dibasic) supplemented with kanamycin (50 $\mu\text{g. mL}^{-1}$). The protein soluble extract was further lysate and purified through immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC) using an ÄKTA purifier system (GE Healthcare) as previously reported by Bonfim and collaborators (data not published, see Chapter 2). Purified samples were analyzed by SDS-PAGE (Supplementary Figure 10) and dynamic light scattering (DLS) in a Malvern ZetaSizer Nano series Nano-ZS (model ZEN3600) instrument (Malvern Zetasizer) previous to enzymatic and crystallization assays. DLS data were collected and analyzed with Zetasizer (7.12) software to evaluate sample homogeneity.

Substrate screening

The enzymatic activity of XagD was evaluated against several synthetic and natural substrates. The reactions with pNP-derived substrates were performed with McIlvaine buffer¹⁸ pH 5.5 containing 5.45 mM of the substrate with a final enzyme concentration of 0.5 mg. mL^{-1} . The reactions, with a final volume of 55 μl , were maintained at 35 °C and 800 rpm for 1h and then interrupted with an equivalent volume of a solution of supersaturated sodium tetraborate. The release of *p*-nitrophenol in each reaction was measured at 400 nm. For the substrate screening on polymeric carbohydrates, the reactions were performed with McIlvaine buffer¹⁸ pH 5.5 containing 2 mg. mL^{-1} of the substrate with a final enzyme concentration of 0.3 mg. mL^{-1} . The reactions, with a final volume of 100 μL , were kept at 35 °C and 1000 rpm for 3h and then interrupted with an equivalent volume of DNS, according to the 3,5-dinitrosalicylic

acid method¹². All the absorbance readings were performed on the Infinite® 200 PRO microplate reader (TECAN).

Biochemical assays

To determine the optimal pH, the substrate galactomannan (carob) was incubated in McIlvaine buffer¹⁸ (pH 3.0 – 8.0) for 30 min at 35 °C and 800 rpm, with a final enzyme concentration equal to 0.3 mg. mL⁻¹ in a final volume of 100 µL. For the determination of the optimal temperature, the tests were carried out in McIlvaine buffer¹⁸ pH 6.0 with a temperature range of 10-60 °C, with a variation of 5 °C between each measurement. All the reactions were interrupted with the addition of 100 µL of DNS.

After defining the optimal reaction conditions, the specific activity of XagD was determined for different substrates. The reactions were prepared for a final volume of 100 µl, containing 2.5 µg of the enzyme, 60 µl of the substrates at 10 mg. mL⁻¹ and 30 µl of McIlvaine buffer¹⁸ pH 6 and incubated at 40 °C and 850 rpm for 25 minutes. To determine the kinetic parameters of XagD, reactions were prepared as described above, but with different concentrations of galactomannans (carob or guar) for a final volume of 100 µl. The enzyme amount and reaction time for enzyme assays were determined based on linearity tests previously performed. Kinetic parameters were determined from substrate saturation curves using the OriginPro 8.1 software. All enzyme assays consisted of at least three independent experiments.

Enzymatic activity evaluated by mass spectrometry

In the cleavage pattern assays with polysaccharides, the substrates galactomannan carob, glucomannan konjac and galactomannan guar (with 21, 28% or 38% of galactose substitutions) were evaluated. The reactions, containing 0.025 mg. mL⁻¹ of enzyme and 2 mg. mL⁻¹ of the substrate, were performed in McIlvaine buffer¹⁸ pH 6 and maintained at 40 °C and 600 rpm for 22 hours. An aliquot of 10 µL was withdrawn at the end of the reaction and added to 40 µL of methanol. For assays with manno-oligosaccharides, cleavage pattern analyzes were performed in reactions containing 0.8 mg. mL⁻¹ of enzyme and 2 mM of substrates in McIlvaine buffer¹⁸ pH 6, at 40°C and 30 min.

For kinetic assays, the reactions contained 5 µL of oligosaccharides in different concentrations (0.25-30 mM for M5 and 0.25-15 mM for M6), 3 µL of McIlvaine buffer¹⁸ at pH 6.0, 1 µL of water and 1 µL of XagD enzyme (in final concentrations of 0.2 mg. mL⁻¹ for M5 and 0.1 mg. mL⁻¹ for M6). After 20 minutes, 20 µL of methanol was added to stop the reactions. The kinetic analyses were performed on a Waters Synapt HDMS, at V mode and

ESI(+) with a spray voltage maintained at 3.0 kV and heated to 130 °C in the source. A total of 15 μL of the quenched reactions and 2 μL of 1 mM xylotriose (used as an internal standard) were added to 134 μL of water and then injected into the mass spectrometer in scan mode (m/z 150-500), with direct infusion at a flow rate of 20 $\mu\text{L} \cdot \text{min}^{-1}$. The internal standard of analyte-like ionization (xylotriose) was used to increase the method's reliability¹⁹. A calibration curve was used to determine the concentration of the enzymatic reaction products. The kinetic parameters of the reactions were determined by non-linear regression analysis (Hill model) of the Michaelis–Menten plot using the software Origin8.1.

Crystallization assays, X-ray data collection and data processing

Crystals of XagD were initially obtained by the sitting-drop vapor-diffusion method in drops containing 0.5 μL of protein solution at 17 mg. mL^{-1} and an equivalent volume of the crystallization solution (0.1 M CHES buffer pH 9, 20% w/v polyethylene glycol (PEG) 4000 and 0.2 M MgCl_2) equilibrated against a reservoir containing 80 μL of the crystallization condition at 18 °C. Crystals were incubated by the soaking method with 0.5 M sodium iodide (NaI) and 10% (v/v) of glycerol cryoprotectant solution for a period of about 60 seconds for the incorporation of heavy metals into the crystalline structure before the data collection. Native crystals of XagD were obtained by the hanging-drop vapor-diffusion method at 18 °C in drops containing 1 μL of protein solution at 17 mg. mL^{-1} and an equivalent volume of the crystallization solution (0.1 M CHES buffer pH 9, 14% w/v polyethylene glycol (PEG) 8000 and 10 mM of mannose) equilibrated against 200 μL of the crystallization condition in 48-well plates. Native crystals were transferred to a cryo-solution containing the crystallization condition, 10% w/v polyethylene glycol (PEG) 400 and 10 mM of mannose before the data collection.

Data were collected under cryogenic conditions (100 K) on the MX2 beamline of the LNLS, equipped with a Pilatus 2M detector (Dectris). Data were indexed, integrated and scaled using the XDS package²⁰. XagD_iodine structure was solved by the Single Anomalous Diffraction method. The positions of heavy atoms and calculation of the initial electron density map were determined using the programs SHELXC/D/E from CCP4i package²¹. Subsequently, the Buccaneer program was used for model building²². The structure of native XagD was solved by the molecular replacement method using the PHASER program²³ and XagD-iodine coordinates as a template. Both structures were refined in initial stages using the REFMAC5 program at CCP4i package²⁴ and then PHENIX refine²⁵ with manual inspections in the program COOT²⁶. The final models were validated using Molprobability²⁷.

Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) data collection was performed at the D01A-SAXS2 beamline at the Brazilian Synchrotron Light Laboratory (LNLS-CNPEM, Campinas, Brazil), using a CCD-Mar165 detector. SAXS measurements of XagD were performed in two different concentrations (5 and 6 mg. mL⁻¹) in 20 mM sodium HEPES buffer with 150 mM NaCl, pH 7.4. Data were integrated using Fit2D²⁸ and processed using the ATSAS package²⁹. The package GNOM³⁰ was used to evaluate the pair-distance distribution functions p(r). Molecular envelopes were calculated from the experimental SAXS data using the program DAMMIN³¹. Averaged models were generated from several runs using DAMAVER suite programs³². The theoretical scattering curve was calculated from the crystallographic model and compared with the experimental SAXS curves using the program CRY SOL³³. The crystallographic structures were fitted into the corresponding SAXS molecular envelopes using the program SUPCOMB³⁴.

Authors' contributions

I.M.B. performed cloning, expression, purification, enzymatic assays and crystallization trials. I.M.B. and M.P.M. performed biophysical analyses. I.M.B., M.P.M., P.O.G. and M.T.M. performed crystallographic studies. R.A.S.P. and F.C.G. performed and analyzed mass spectrometry experiments. I.M.B. and M.P.M. analyzed the results and wrote the manuscript. P.O.G. and M.T.M. coordinated the project, analyzed the results and wrote the manuscript.

References

1. Zhang, Y. *et al.* Biochemical and structural characterization of the intracellular mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel glycoside hydrolase family belonging to clan GH-A. *J. Biol. Chem.* **283**, 31551–31558 (2008).
2. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
3. Couturier, M. *et al.* Functional exploration of the glycoside hydrolase family GH113. *PLoS One* **17**, e0267509 (2022).
4. Sun, D. *et al.* Reshaping the binding channel of a novel GH113 family β -mannanase from *Paenibacillus cineris* (PcMan113) for enhanced activity. *Bioresour. Bioprocess.* **9**, (2022).
5. La Rosa, S. L. *et al.* The human gut Firmicute *Roseburia intestinalis* is a primary degrader of dietary β -mannans. *Nat. Commun.* **10**, 1–14 (2019).
6. You, X. *et al.* Structural insights into the catalytic mechanism of a novel glycoside hydrolase family 113 -1,4-mannanase from *Amphibacillus xylanus*. *J. Biol. Chem.* **293**, 11746–11757

- (2018).
7. Lindstad, L. J. *et al.* Human gut *Faecalibacterium prausnitzii* deploys a highly efficient conserved system to cross-feed on β -mannan derived oligosaccharides. *MBio* **12**, (2021).
 8. Xia, W. *et al.* A Novel Glycoside Hydrolase Family 113 Endo- β -1,4-Mannanase from *Alicyclobacillus* sp. Strain A4 and Insight into the Substrate Recognition and Catalytic Mechanism of This Family. *Appl. Environ. Microbiol.* **82**, 2718–2727 (2016).
 9. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
 10. Liu, W. *et al.* Functional and structural investigation of a novel β -mannanase BaMan113A from *Bacillus* sp. N16-5. *Int. J. Biol. Macromol.* **182**, 899–909 (2021).
 11. Tao, F., Swarup, S. & Zhang, L. H. Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for *Xanthomonas campestris* biofilm formation. *Environ. Microbiol.* **12**, 3159–3170 (2010).
 12. Miller, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **31**, 426–428 (1959).
 13. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
 14. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
 15. Mazur, O. & Zimmer, J. Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. *J. Biol. Chem.* **286**, 17601–17606 (2011).
 16. Itoh, Y. *et al.* Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly- β -1,6-N-acetyl-D-glucosamine. *J. Bacteriol.* **190**, 3670–3680 (2008).
 17. Bakkevig, K. *et al.* Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *J. Bacteriol.* **187**, 8375–8384 (2005).
 18. McIlvaine, T. C. A buffer solution for colorimetric comparison. *JOURNAL OF BIOLOGICAL CHEMISTRY* vol. v.49:183-1 (1921).
 19. Ge, X., Sirich, T. L., Beyer, M. K., Desaire, H. & Leary, J. A. A Strategy for the Determination of Enzyme Kinetics Using Electrospray Ionization with an Ion Trap Mass Spectrometer. *Anal. Chem.* **73**, 5078–5082 (2001).
 20. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).
 21. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D* **66**, 479–485 (2010).
 22. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D* **62**, 1002–1011 (2006).
 23. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
 24. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D* **67**, 355–367 (2011).
 25. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).
 26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).

27. Davis, I. W. *et al.* MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, 375–383 (2007).
28. Hammersley, A. P. *et al.* Calibration and Application of an X-ray Image Intensifier/ Charge-Coupled Device Detector for Monochromatic Macromolecular Crystallography. *J. Synchrotron Radiat.* **4**, 67–77 (1997).
29. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
30. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
31. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
32. Volkov, V. V & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
33. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
34. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).

Supplementary Information

Structural and functional characterization of a GH113 enzyme from a phytopathogenic Proteobacteria reveal molecular adaptations for a role in EPS cleavage

Isabela Mendes Bonfim^{1,2}, Marcele Pandeló Martins¹, Mariane Noronha Domingues¹, Renan Augusto Siqueira Pirolla^{1,3,4}, Fabio Cesar Gozzo⁴, Priscila Oliveira de Giuseppe^{1,*} and Mário Tyago Murakami^{1,*}

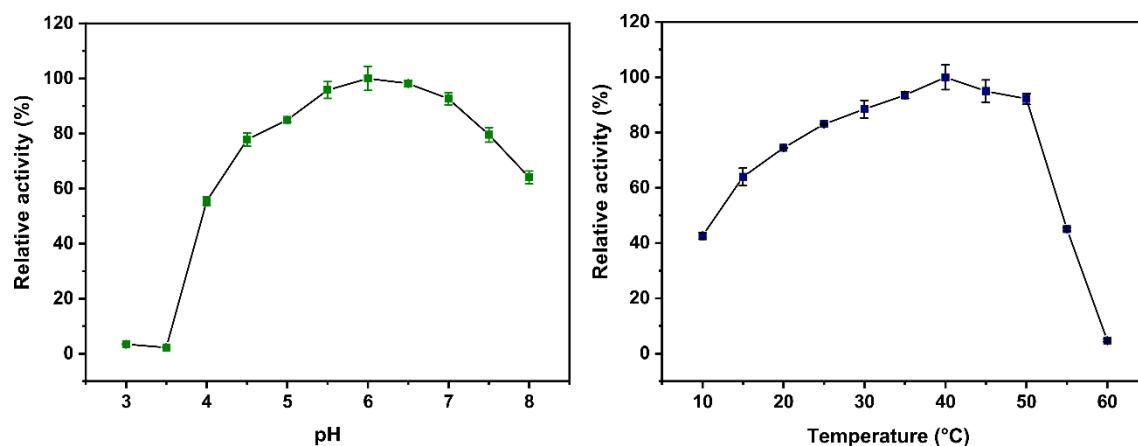
¹Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPem), Zip Code 13083-970, Campinas, Sao Paulo, Brazil.

²Graduate Program in Functional and Molecular Biology, Institute of Biology, University of Campinas, Campinas, São Paulo, Brazil

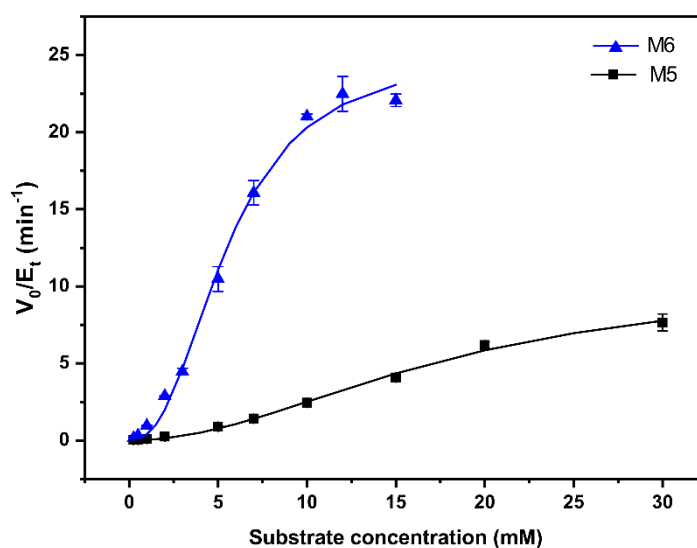
³ Current address: Waters Technologies Brasil, Barueri, São Paulo, Brazil

⁴ Institute of Chemistry, University of Campinas, Campinas, São Paulo, Brazil

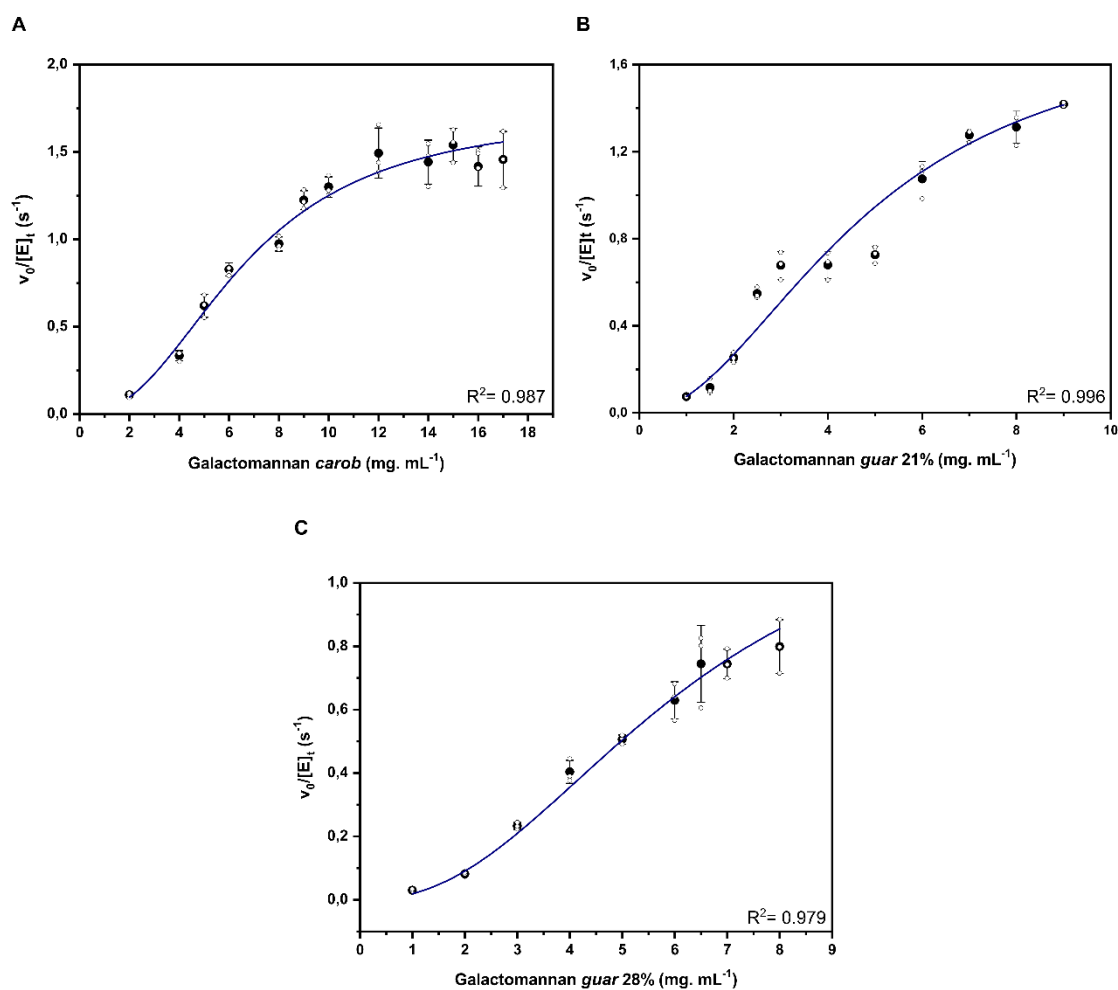
*Correspondence to: priscila.giuseppe@lnbr.cnpem.br and mario.murakami@lnbr.cnpem.br



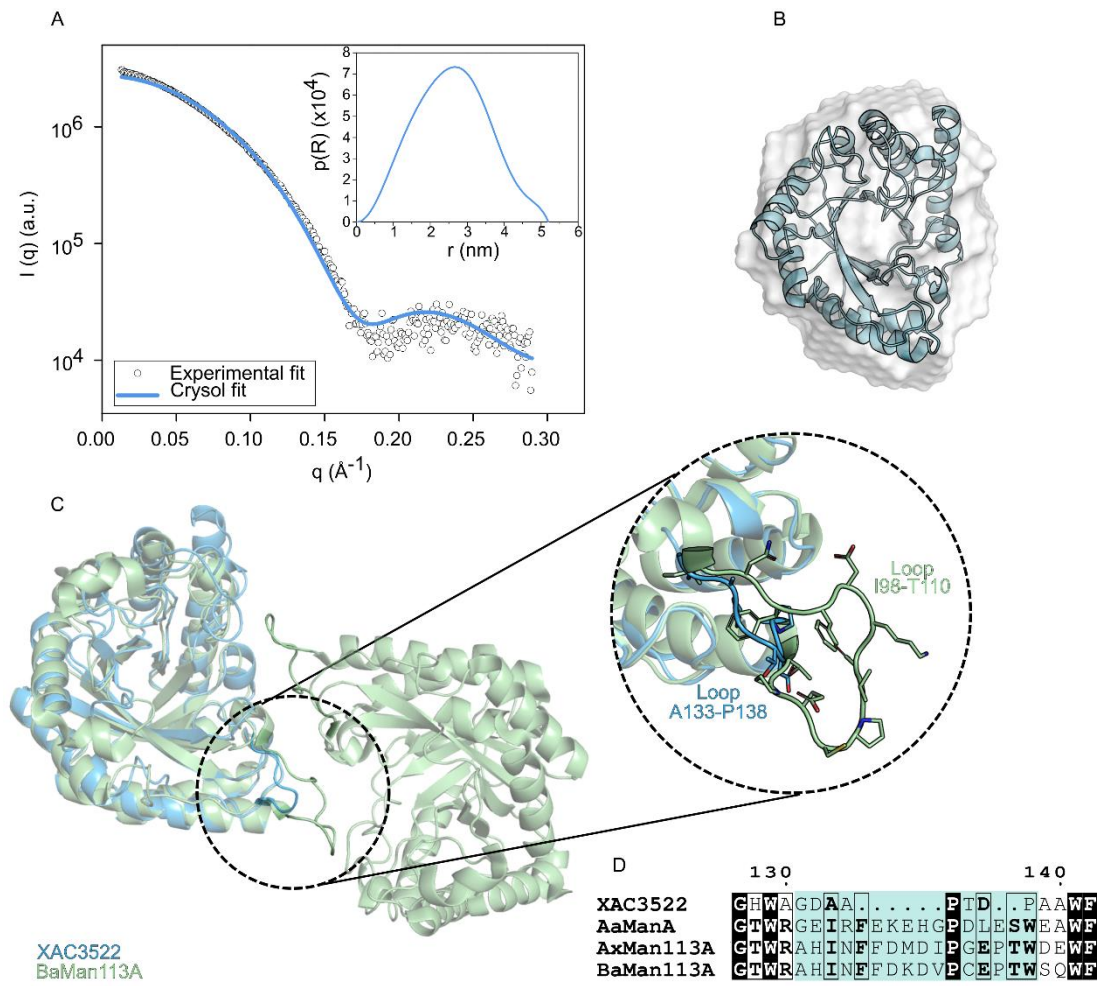
Supplementary Figure 1. pH and Temperature dependence curves of XagD activity. Data are shown as mean \pm SD from three independent experiments (n=3).



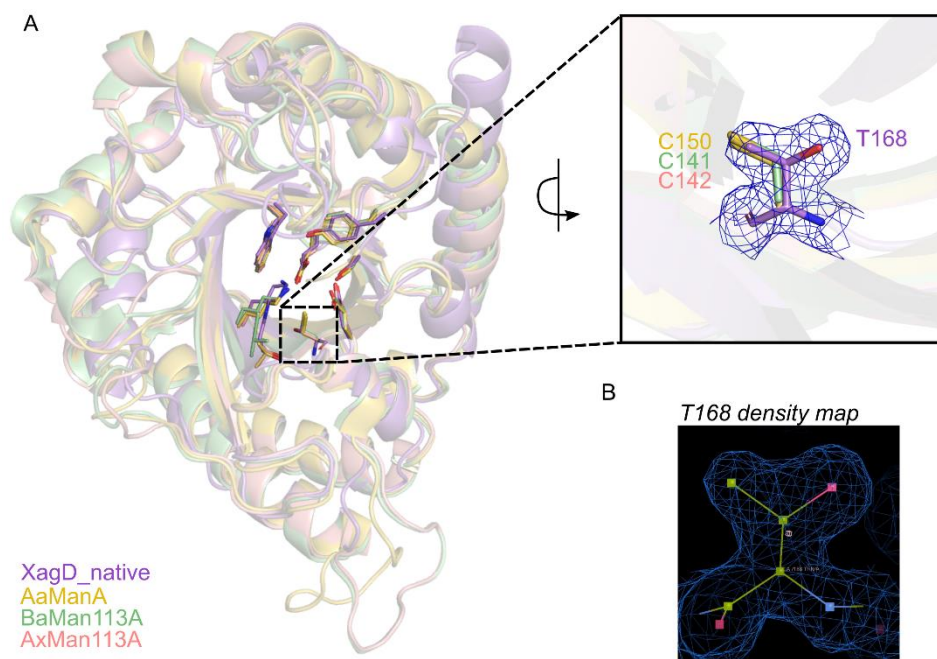
Supplementary Figure 2. Saturation curve of XagD on the oligosaccharides mannopentaose (M5) and mannohexaose (M6). The k_{cat} and K_m values were calculated using Hill's model. Data are shown as mean \pm SD from three independent experiments (n=3).



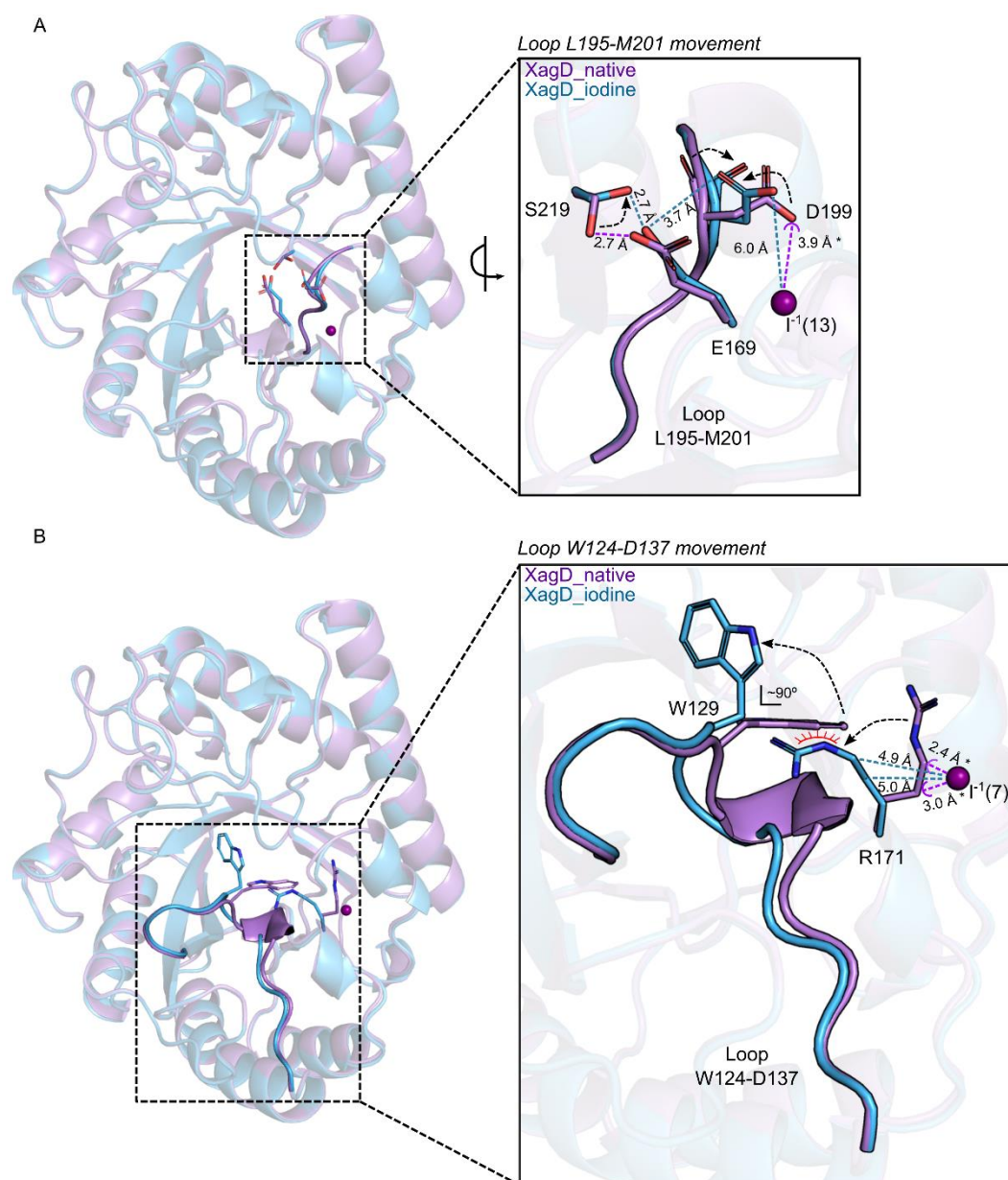
Supplementary Figure 3. Saturation curves of XagD on different polysaccharides. A) Galactomannan carob, **B)** galactomannan guar (21%) and **C)** galactomannan guar (28%). From these curves, the k_{cat} and K_m values were calculated using Hill's model. Data are shown as mean \pm SD from three independent experiments ($n=3$). The adjusted R^2 of the curves is indicated in each graph.



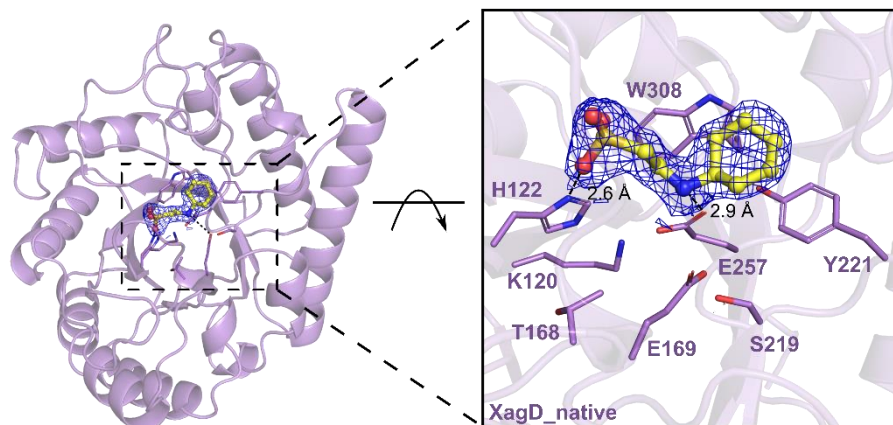
Supplementary Figure 4. Analysis of the oligomeric state of XagD. **A)** SAXS experimental scattering curve (black circles) and theoretical scattering curve (blue line) calculated using CRY SOL¹ based on the XagD crystallographic structure. The distance distribution curve ($P(r)$) generated by GNOM² exhibits a D_{max} of 5.2 nm. **B)** XagD envelope calculated from SAXS data, superimposed onto the XagD crystallographic structure (blue cartoon). **C)** Structure comparison between the monomer XagD (blue) and the dimer *BaMan113A* (pale green, rmsd 1.115 Å). The loop that stabilizes the quaternary structure of *BaMan113A* is represented in the zoomed figure, highlighting the length difference between the loops in XagD and *BaMan113A*. **D)** Multiple sequence alignment of XagD and GH113 structures solved (*AaManA*, *AxMan113A*, and *BaMan113A*). The region highlighted in blue corresponds to loop A133-P138, which is shortened in XagD.



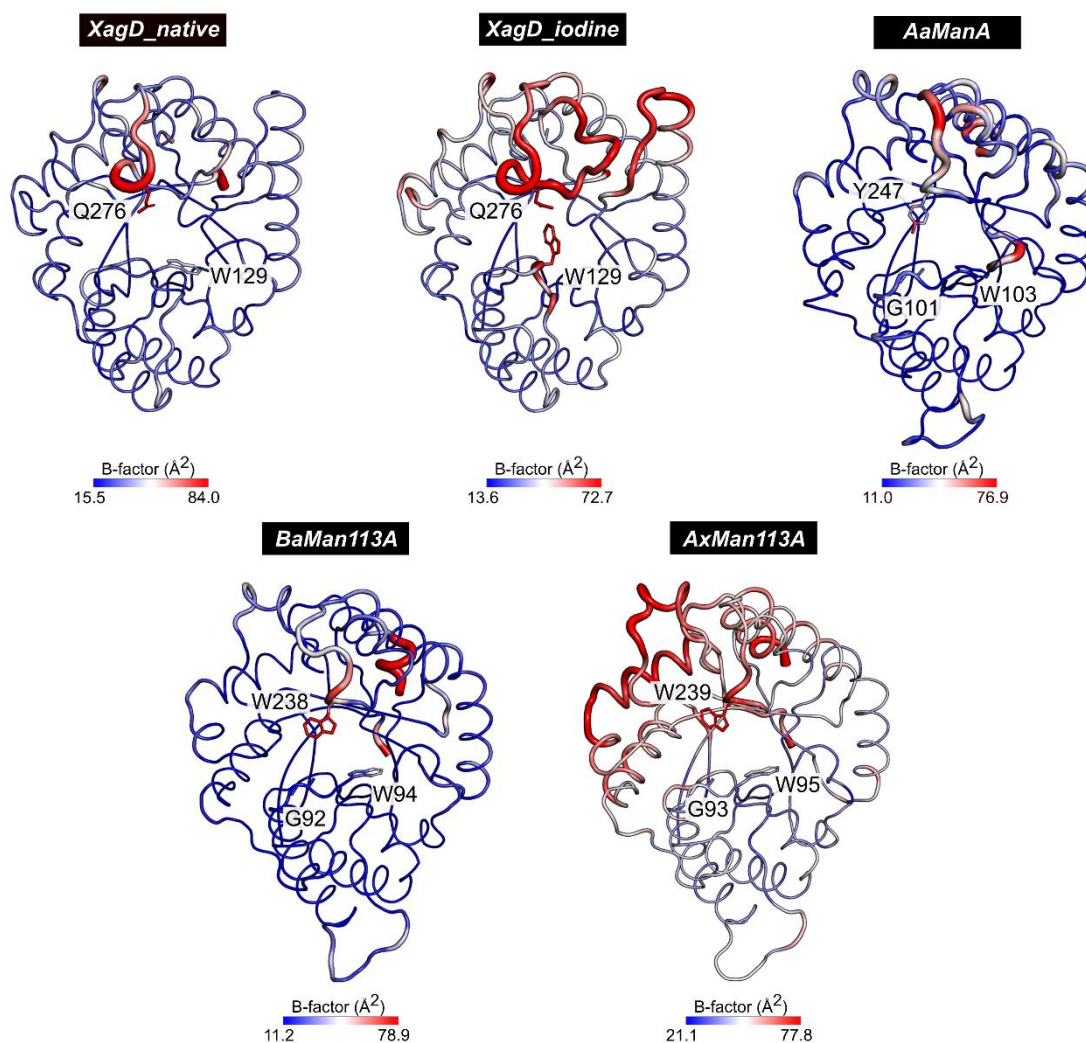
Supplementary Figure 5. A) Structural comparison of XagD_native, AaManA (PDB code 4CD8, rmsd 1.12 Å, yellow), BaMan113A (PDB code 7DV7, rmsd 1.12 Å, green), and AxMan113A (PDB code 5YLH, rmsd 1.30 Å, pink), evidencing the electron density map of T168 in XagD_native. The electron density map (2Fo-Fc) around the T168 is represented at the 1.0 σ contour level. **B)** View of the electron density map of T168 in XagD_native using the COOT software³.



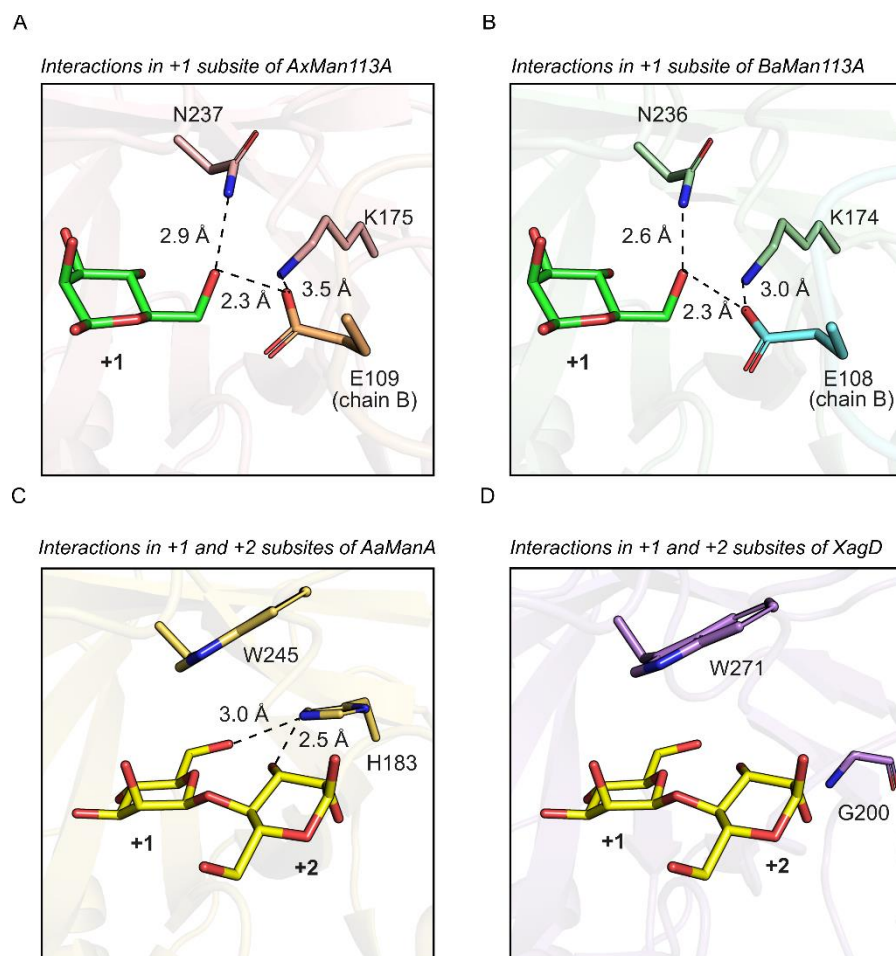
Supplementary Figure 6. Loops movements in XagD caused by iodine ions binding. **A)** Loop L195-M201 movement. The dashed arrows indicate the residue movement from XagD_native to XagD_iodine data due to I13 binding. The purple distance represented by a semi-circle marker (3.9 Å *) is the theoretical distance from I13 to D199 from XagD_native. **B)** Loop W124-D137 movement. The dashed arrows indicate the residues movements from XagD_native to XagD_iodine data due to I7 binding. The purple distances represented by semi-circle markers (2.4 Å and 3.0 Å *) are the theoretical distance from I7 to R171 from XagD_native. The red semi-circle indicates the steric hindrance between XagD_iodine-R171 and XagD_native-W129 side chains, causing the W129 movement around 90 ° in XagD_iodine.



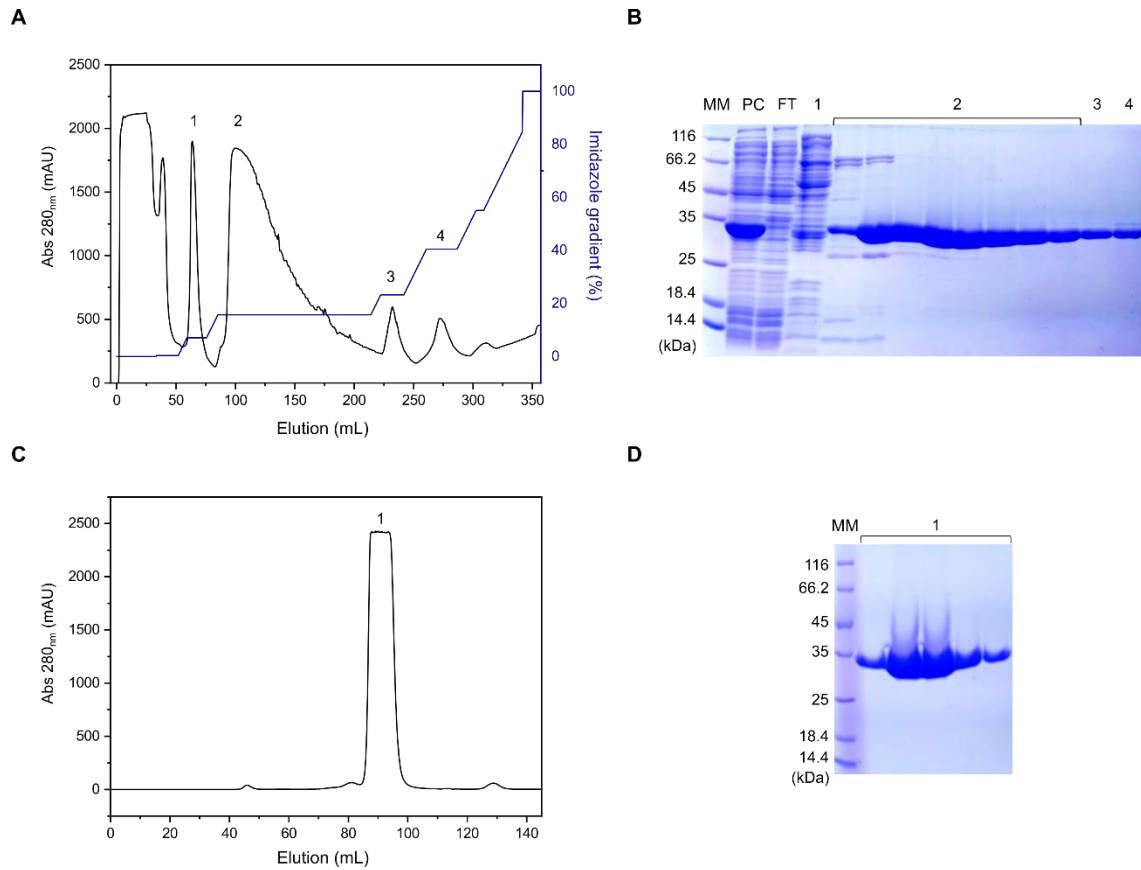
Supplementary Figure 7. Interaction of XagD residues with a CHES molecule in the active site. The electron density map (2Fo-Fc) around the CHES is represented at 1.0 σ contour level. Dashed lines represent interatomic distances.



Supplementary Figure 8. Cartoon representation of GH113 enzymes crystallographic structure colored by B-factor. The color gradient indicates the protein regions with the lowest (blue), medium (white) and highest (red) values of B-factor. The residues involved in lid-formation mechanism are highlighted in XagD_native, XagD_iodine, AaManA (PDB code 4CD8), BaMan113A (PDB code 7DV7), and AxMan113A (PDB code 5YLH).



Supplementary Figure 9. Comparisons of +1 and +2 subsites interactions with substrate in monomeric and dimeric GH113 enzymes. Interactions of **A)** AxMan113A (PDB code 5YLK, chain A in pink and chain B in orange), **B)** BaMan113A (PDB code 7DV7, chain A in light green and chain B in light blue), **C)** AaManA (PDB code 4CD8, yellow) and **D)** XagD_native (purple) with substrates in positive subsites. Mannose (green) and mannobiose (yellow) coordinates are from the crystal structures of enzyme-product complexes of AxMan113A (PDB code 5YLK) and AaManA (PDB code 4CD7), respectively.



Supplementary Figure 10. Purification profile of XagD. **A)** Immobilized metal affinity chromatography (IMAC) chromatogram and **B)** SDS-PAGE (13%) analysis of IMAC samples. **C)** Size-exclusion chromatography (SEC) chromatogram; **D)** SDS-PAGE (13%) analysis of SEC samples. MM = molecular marker and the following fractions corresponding to the peak observed in the chromatogram of panel C. MM: molecular marker; PC: pre-column; FT: flow-through; Numbers correspond to the elution peaks from which samples were selected and analyzed by SDS-PAGE.

Supplementary Table 1. List of synthetic substrates used for activity detection screening of XagD enzyme. The results are qualitative and indicate the presence (+) or absence (-) of detected enzymatic activity. All substrates were provided by Megazyme.

Substrate	Enzymatic activity
4-Nitrophenyl- β -D-mannopyranoside	-
4-Nitrophenyl- α -D-mannopyranoside	-
4-Nitrophenyl- β -D-glucopyranoside	-
4-Nitrophenyl- α -D-xylopyranoside	-
4-Nitrophenyl- β -D-xylopyranoside	-
4-Nitrophenyl- β -D-galactopyranoside	-
4-Nitrophenyl- α -D-galactopyranoside	-
4-Nitrophenyl- α -L-arabinofuranoside	-
4-Nitrophenyl- β -D-cellobioside	-

Supplementary Table 2. List of polymeric substrates used for activity detection screening of XagD enzyme. The results are qualitative and indicate the relative enzymatic activity through the 3,5-dinitrosalicylic acid method⁴.

Polymeric substrate	Enzymatic activity
Arabinnan ¹	16,7
Arabinogalactan ¹	11,1
Arabinoxylan (rye flour) ¹	22,7
Arabinoxylan (wheat flour) ¹	23,0
CMC 4M ²	20,0
CM curdlan ¹	19,1
Galactan (lupin) ¹	26,4
Galactan (potato) ¹	21,4
Galactomannan (carob, low viscosity)¹	77,4
Galactomannan (guar, 28%)¹	88,3
Glucomannan (konjac, low viscosity)¹	100,0
Laminarin (from <i>Laminaria digitata</i>) ²	31,4
Pectin (from citrus peel) ²	13,5
Xanthan gum ²	14,5
Xylan (from beechwood) ²	17,5
Xyloglucan (amyloid) ¹	28,3
Xyloglucan (tamarind seed) ¹	25,5
Beta-glucan (barley) ¹	25,4
Chitosan ¹	21,5
Mannan (borohydride reduced) ¹	17,0
Mannan (ivory nut) ¹	37,8
Rhamnogalacturonan (potato) ¹	25,2
Starch (soluble, potato) ²	9,2

¹Megazyme

²Sigma-Aldrich

Supplementary Table 3. Comparison of the parameter of specific activity of GH113 enzymes over different heteromannans. When available, the different content of galactose ramifications in galactomannan guar gum were indicated in parenthesis. LBG, locust bean gum; KGM, glucomannan konjac; NT, not tested. ND, not detected.

Protein	LBG (U. mg ⁻¹)	Guar Gum (U. mg ⁻¹)	KGM (U. mg ⁻¹)	Reference
		3.9 ± 0.06 (~21%)		
XagD	1.17 ± 0.1	1.5 ± 0.07 (28%)	1.44 ± 0.03	This work
		ND (38%)		
<i>AaManA</i> [#]	724.4 ± 20	79.4 ± 1.2 (38%)	1055 ± 26.2	5
		391.3 ± 4.5 (21%)		
Man113A	298.8 ± 6.4	302.8 ± 2.5 (28%) 36.2 ± 2.2 (34%) 34.9 ± 4.9 (38%)	370.4 ± 11.8	6
<i>AxMan113A</i> WT	22.4 ± 0.2 (^{>M1})	1.2 ± 0.02	54.4 ± 0.66	7
<i>BaMan113A</i> WT	0.94 ± 0.03	NT	1.04 ± 0.08	8
<i>PcMan113</i> WT	31.7 ± 0.1	2.1 ± 0.07	68.2 ± 0.36	9

Supplementary Table 4. X-ray diffraction data collection and refinement statistics. Values in parenthesis represent the higher resolution shell.

Data collection	XagD_iodine	XagD_native
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	58.0, 62.52, 74.76	62.27, 77.15, 90.73
α , β , γ (°)	90	90
Resolution (Å)	47.96 - 1.79 (1.86 - 1.79)	19.59 - 1.78 (1.85 - 1.78)
Total reflections	264,711 (20,816)	340,696 (19,179)
Unique reflections	48,664 (2,558)	41,713 (3,719)
R _{meas}	0.104 (1.43)	0.086 (0.91)
<i>I</i> / σ <i>I</i>	17.29 (1.57)	16.21 (1.78)
Completeness (%)	99.89 (99.88)	98.02 (88.80)
Redundancy	10.2 (8.1)	8.2 (5.2)
CC $\frac{1}{2}$ (%) [§]	99.9 (73.3)	99.9 (81.3)
Refinement		
Resolution (Å)	29.0 - 1.79	19.59 - 1.78
Number of reflections	25,859	41,683
Number of protein residues	296	297
R _{work} /R _{free}	0.204 (0.336)/ 0.245 (0.350)	0.172 (0.296)/ 0.196 (0.313)
B-factor (Å ²)		
Macromolecules	32.35	28.15
Ligands	46.55	53.23
Water	38.74	39.93
Root mean square deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.23	1.23
Ramachandran plot		
Favored (%)	98.98	98.98
Outliers (%)	0	0

[§]CC_{1/2}: correlation between intensities from random half-datasets¹⁰

Supplementary Table 5. Description of lid conformation on GH113 structures deposited at the PDB and XagD.

Protein	Lid	Ligand	PDB code	Reference
<i>AaManA</i>	Opened	-	3CIV	5
<i>AaManA</i>	Closed	ManIFG	4CD6	11
<i>AaManA</i>	Closed	ManIFG+M2	4CD7	11
<i>AaManA</i>	Opened	ManMIm	4CD8	11
<i>AxMan113A</i>	Opened	-	5YLH	7
<i>AxMan113A</i>	Opened	M4	5YLI	7
<i>AxMan113A</i>	Closed	M2	5YLK	7
<i>AxMan113A</i>	Opened	M6 ¹	5YLL	7
<i>AxMan113A</i>	Opened	M3 ²	5Z4T	7
<i>BaMan113A</i>	Opened	-	7DV7	8
<i>BaMan113A</i>	Opened	M2	7DVJ	8
<i>BaMan113A</i>	Opened	-	7DVZ	8
<i>BaMan113A</i>	Closed	M2	7DW8	8
<i>BaMan113A</i>	Closed	M3	7DWA	8
XagD_native	Opened	CHES	-	This work
XagD_iodine	Closed	CHES	-	This work

¹ M3 molecule is observed in the active site

² M2 molecule is observed in the active site

Supplementary Table 6. Sequence-based prediction of subcellular localization of the XagD protein. TMH = transmembrane helix, PER = periplasm.^a SPII signal predicts for lipoproteins, i.e membrane-tethered proteins via N-terminal lipidation. ^b Position +2 different from D (aspartic acid) after a SPII cleavage site indicates outer membrane anchoring¹².

<i>Locus</i>	Description	Length	SignalP 6.0 ¹³	TMHMM ¹⁴	CELLO v.2.5 ¹⁵	SOSUI/GramN ¹⁶	Consensus	Criterium
XAC3522	Endo- β -1,4-mannanase	337 amino acids	Sec/SPII ^a (1-35)/ +2=N	1 TMH (20-42)	PER	PER	OM	N-terminal lipid anchor/ +2 \neq D ^b

References

1. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
2. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
3. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).
4. Miller, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **31**, 426–428 (1959).
5. Zhang, Y. *et al.* Biochemical and structural characterization of the intracellular mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel glycoside hydrolase family belonging to clan GH-A. *J. Biol. Chem.* **283**, 31551–31558 (2008).
6. Xia, W. *et al.* A Novel Glycoside Hydrolase Family 113 Endo- β -1,4-Mannanase from *Alicyclobacillus* sp. Strain A4 and Insight into the Substrate Recognition and Catalytic Mechanism of This Family. *Appl. Environ. Microbiol.* **82**, 2718–2727 (2016).
7. You, X. *et al.* Structural insights into the catalytic mechanism of a novel glycoside hydrolase family 113 -1,4-mannanase from *Amphibacillus xylanus*. *J. Biol. Chem.* **293**, 11746–11757 (2018).
8. Liu, W. *et al.* Functional and structural investigation of a novel β -mannanase BaMan113A from *Bacillus* sp. N16-5. *Int. J. Biol. Macromol.* **182**, 899–909 (2021).
9. Sun, D. *et al.* Reshaping the binding channel of a novel GH113 family β -mannanase from *Paenibacillus cineris* (PcMan113) for enhanced activity. *Bioresour. Bioprocess.* **9**, (2022).
10. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* (80). **336**, 1030–1033 (2012).
11. Williams, R. J. *et al.* Combined inhibitor free-energy landscape and structural analysis reports on the mannosidase conformational coordinate. *Angew. Chemie - Int. Ed.* **53**, 1087–1091 (2014).
12. Seydel, A., Gounon, P. & Pugsley, A. P. Testing the ‘+2 rule’ for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol. Microbiol.* **34**, 810–821 (1999).
13. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-021-01156-3.
14. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
15. Yu, C.-S., Chen, Y.-C., Lu, C.-H. & Hwang, J.-K. Prediction of protein subcellular localization. *Proteins Struct. Funct. Bioinforma.* **64**, 643–651 (2006).
16. Imai, K. *et al.* SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. *Bioinformation* **2**, 417–421 (2008).

Chapter 5

Concluding remarks

Considering the different roles that carbohydrates can display as a carbon source, as signaling molecules or as structural components, this work shed light on how complex plant carbohydrates can affect the transcriptional response of the bacterium *X. citri*, the causal agent of citrus canker, guiding this phytopathogen across the host's environment and influencing its infection strategies to evade the host's immune response and uptake nutrients.

In Chapter 2, using a genome-wide transcriptional approach, we expanded the knowledge of new polysaccharides utilization systems used by *X. citri* to sense, break down and uptake different carbohydrates, evidencing the capacity of this bacterium to differentiate the chemical composition, glycosidic linkages and decorations of each polysaccharide, which ultimately modulates the physiological responses of this bacterium. This was observed, for example, in the presence of cellobiose and starch, both polysaccharides are composed of glucose residues but are connected by β - and α - glycoside linkages, respectively, and were capable of triggering opposite transcriptional responses related to bacterial motility and sulfur metabolism.

In Chapter 3, our transcriptional studies were instrumental in revealing the importance of xyloglucan depolymerization and uptake for triggering the activation of key virulence factors, including the type III secretion system (T3SS), several type III effectors (T3E) and the production of xanthan gum, together with the downregulation of chemotaxis and flagellar-dependent motility. In plants, xyloglucan is mainly present in the primary cell wall, which is less lignified and more accessible for enzymatic degradation. Together with our results, this suggests that xyloglucan might display a significant role as a barrier for *Xanthomonas* phytopathogens, which agrees with the activation of the XyGUL molecular machinery to depolymerize this carbohydrate concomitant with the activation of T3SS, allowing the disruption of the plant cell wall to translocate T3E that will modulate the host metabolism and allow a successful development of citrus canker infection.

In Chapter 4, we used biochemical and structural analyses to evaluate the glycoside hydrolase belonging to the Xag-EPS biosynthesis cluster in *X. citri*, XagD (XAC3522). Our results bring novel insights into the underexplored family GH113 and provide clues about the putative structure of Xag-EPS and for the possible biological roles of XagD.

In summary, this study reveals how different approaches, including transcriptomics and biochemical and structural characterization of enzymes, can be combined to explore the influence of carbohydrates on bacterial metabolism and to reveal new PULs or enzymes related to the degradation of complex carbohydrates present in nature. This work advances the knowledge of polysaccharide utilization systems in Proteobacteria and shows how plant polysaccharides can influence the host-pathogen interactions.

Anexo I. Termo de autorização do comitê de biossegurança

Uso exclusivo da CIBio:

Número de projeto / processo: 2019-13

Formulário de encaminhamento de projetos de pesquisa com OGMs para análise da CIBio - CNPEM

1. Título do projeto: Estudos estruturais, funcionais e desenho racional de hidrolases glicosídicas que atuam na despolimerização de polissacarídeos da biomassa

2. Pesquisador responsável: Mário Tyago Murakami

3. Experimentador(es):

Camila Ramos dos Santos, Priscila Oliveira de Giuseppe, Mariane Noronha Domingues, Atilio Tomazini Júnior, Mariana Abrahão Bueno de Moraes, Plínio Salmazo Vieira, Lucélia Cabral, Thamy Livia Correa, Pedro Avellar Cabral, Rodrigues da Costa, Rosa Lorizolla Cordeiro, Isabela Mendes Bonfim, Mariana Chinaglia, José Alberto Diogo, Marcele Pandelo Martins, Liliane Pires Andrade, Ellen Karen Barreto Roman, Jéssica Batista de Lima Correa, Beatriz Paiva de Souza, Daniel Fernando Coletta da Silva, Augusto Rodrigues Lima

Nível do treinamento do experimentador:

[EKBR, JBLC, BPS]-Iniciação científica, [] -mestrado, [JAD, MPM, LPA]-doutorado, [PACRC, RLC, IMB, MC]-doutorado direto, [CRS, POG, MND, ATJ, MABM, PSV, LC, TLC]-pós-doutorado, [] -nível técnico, [DFCS, ARL]-outro, especifique: estágio

4. Unidade operativa: [] LNLS [] LNNano [X] LNBR [] LNBio

5. Maior Classe de risco de OGM deste projeto: [X] Risco I [] Risco II [] Risco III [] Risco IV

6. O projeto é confidencial? [X] não [] sim

7. No caso de projeto confidencial, o título do projeto pode constar em lista aberta no CNPEM? [] não [X] sim

8. Qual é o objetivo do projeto? Neste projeto visamos explorar e entender, ao nível mecanístico, a diversidade funcional das GHs de nichos ainda pouco explorados como o fitopatógeno Xac e de subfamílias inéditas. Para tanto, utilizaremos uma abordagem multidisciplinar que integrará estudos evolutivos, bioquímicos e estruturais, visando fornecer resultados que potencialmente revelarão novas estratégias para a despolimerização da parede celular vegetal e gerarão dados instrumentais para o uso racional e aperfeiçoado de GHs para fins industriais.

9. Informe um número e nome para cada OGM, organismo receptor, organismo doador, o transgene e classe de risco do OGM.

(1) Ecoli-pET28a-XAC1770, Escherichia coli, Xanthomonas citri, XAC1770, risco I, (2) Ecoli-pET28a-XAC3072, Escherichia coli, Xanthomonas citri, XAC3072, risco I, (3) Ecoli-pET28a-XAC1796, Escherichia coli, Xanthomonas citri, XAC1796, risco I, (4) Ecoli-pET28a-XAC3522, Escherichia coli, Xanthomonas citri, XAC3522, risco I.

10. Descreva brevemente a função dos transgenes de cada OGM: (1,2,3,4) hidrolases glicosídicas que atuam sobre a parede celular vegetal.

11. Algum OGM produz proteína tóxica, oncogênica ou pode gerar produtos deletérios para saúde humana, animal ou meio ambiente? Não

12. Algum OGM é agente patogênico esporulante? [X] Não [] Sim: _____

13. Algum OGM é agente patogênico e pode se propagar pelo ar? [X] Não [] Sim: _____

14. Algum transgene confere infectividade ou patogenicidade para os OGMs? Descreva. Todos os OGMs apresentam resistência ao antibiótico canamicina. Nenhum transgene confere infectividade ou patogenicidade ao organismo receptor.

15. Com relação aos cuidados preventivos associados a manipulação dos organismos, será necessária alguma avaliação médica periódica para experimentadores? [X] Não [] Sim. **Que tipo de avaliação?** (Ex: consulta com médico, exames laboratoriais etc...) **Qual periodicidade? Onde será realizada esta avaliação?**

Uso exclusivo da CIBio:

Número de projeto / processo: 2019-13

Formulário de encaminhamento de projetos de pesquisa com OGMs para análise da CIBio - CNPEM

16. Com relação aos cuidados preventivos associados a manipulação dos organismos, será necessária alguma vacinação preventiva para experimentadores? ☒ Não ☐ Sim. Qual periodicidade? Onde será realizada esta vacinação?

17. No caso de uma eventual contaminação com organismos patogênicos ou toxinas, descreva medidas emergenciais para tratamento de pessoas envolvidas, descontaminação de equipamentos, instalações e meio ambiente. Não há necessidade

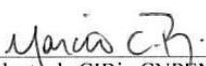
18. Projetos que façam uso de organismos ou genes associados ao patrimônio genético brasileiro precisam de cadastro na plataforma SISGEN (www.sisgen.gov.br). É de total responsabilidade do pesquisador responsável esse cadastramento e cumprimento da legislação. O projeto envolve manipulação, transferência, modificação, armazenamento, coleta de Organismos e derivados relativos ao patrimônio genético brasileiro? ☒ SIM, ☐ Não. No caso de responder sim, mencionar a seguir quais os códigos de acesso do cadastro no SISGEN: A3C94FA.

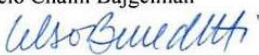
O pesquisador principal tem conhecimento de que conforme a RDC 50 de 21/02/2002 da Anvisa, é responsável por determinar a classificação de riscos de seu projeto, assim como determinar EPIs e medidas de segurança necessárias para prevenir a contaminação de experimentadores, equipamentos, instalações, terceiros e meio ambiente. O pesquisador responsável também precisará providenciar rotina para realização de exames médicos e laboratoriais para sua equipe, bem como vacinações quando aplicável. Todos os experimentadores envolvidos devem ser supervisionados pelo pesquisador principal, que é o responsável pelo treinamento de biossegurança adequado às suas necessidades para a manipulação, armazenamento, descarte e transporte de OGMs, atendendo a legislação e normativas preconizadas pela CTNBio, Anvisa e outros órgãos e agências regulamentadoras e fiscalizadoras.

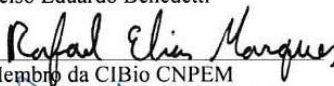
A CIBio analisou este projeto em reunião realizada no dia: 27/6/2019.

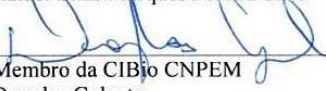
Parecer final: ☒ -projeto aprovado, ☐ -projeto recusado, ☐ -projeto com deficiências.

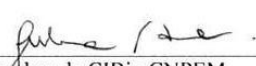
comentários da CIBio:

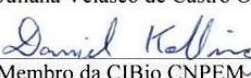

Presidente da CIBio CNPEM
Marcio Chaim Bajgelman

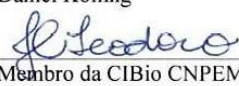

Membro da CIBio CNPEM
Celso Eduardo Benedetti

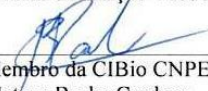

Membro da CIBio CNPEM
Rafael Elias Marques Pereira Silva


Membro da CIBio CNPEM
Douglas Galante


Membro da CIBio CNPEM
Juliana Velasco de Castro Oliveira


Membro da CIBio CNPEM
Daniel Kolling


Membro da CIBio CNPEM
Juliana Conceição Teodoro


Membro da CIBio CNPEM
Mateus Borba Cardoso