Article

Resolving the metabolism of monolignols and other lignin-related aromatic compounds in *Xanthomonas citri*

Damaris B. Martim [1,2], Anna J. V. C. Brilhante [1,2], Augusto R. Lima [1,2], Douglas A. A. Paixão[2], Joaquim Martins-Junior [2], Fernanda M. Kashiwagi [2], Lucia D. Wolf[2], Mariany S. Costa[2], Fabrícia F. Menezes[2], Rafaela Prata [2], Matheus C. Gazolla [2], Juliana A. Aricetti[2], Gabriela F. Persinoti [2], George J. M. Rocha [2] & Priscila O. Giuseppe [1,2] ✉

Lignin, a major plant cell wall component, has an important role in plant-defense mechanisms against pathogens and is a promising renewable carbon source to produce bio-based chemicals. However, our understanding of microbial metabolism is incomplete regarding certain lignin-related compounds like *p*-coumaryl and sinapyl alcohols. Here, we reveal peripheral pathways for the catabolism of the three main lignin precursors (*p*-coumaryl, coniferyl, and sinapyl alcohols) in the plant pathogen *Xanthomonas citri*. Our study demonstrates all the necessary enzymatic steps for funneling these monolignols into the tricarboxylic acid cycle, concurrently uncovering aryl aldehyde reductases that likely protect the pathogen from aldehydes toxicity. It also shows that lignin-related aromatic compounds activate transcriptional responses related to chemotaxis and flagellar-dependent motility, which might play an important role during plant infection. Together our findings provide foundational knowledge to support biotechnological advances for both plant diseases treatments and conversion of lignin-derived compounds into bio-based chemicals.

The bioconversion of aromatic compounds has a central role in carbon cycling[1], plant-pathogen interactions[2–4] and detoxification of organic pollutants[5]. This process typically starts in upper pathways (also named peripheral pathways), which converge the structural diversity of aromatic compounds to fewer intermediate metabolites, which are further funneled to central carbon metabolites through a narrower range of lower pathways[6,7]. During this convergent process, industrially relevant molecules are formed. This offers opportunities for engineering microbial chassis to produce chemicals from complex mixtures of aromatic compounds derived from abundant wastes such as lignin and mixed-plastics[8,9].

Hundreds of microorganisms across diverse phyla are known to have the potential to metabolize lignin-related monomers, according to the eLignin database[7]. However, the metabolic pathways of only a few species have been characterized so far[7,10]. Lignin-related monomers include the three main lignin precursors (*p*-coumaryl, coniferyl and sinapyl alcohols) and their respective *p*-hydroxyphenyl (H), guaiacyl (G) and sinapyl (S) derivatives. Although pathways for the catabolism of several lignin-related monomers have been described in the literature, information on upper pathways for some of them, such as *p*-coumaryl alcohol, sinapyl alcohol and sinapaldehyde, is still lacking[11–13].

[1]Graduate Program in Genetics and Molecular Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil. [2]Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil. ✉e-mail: priscila.giuseppe@lnbr.cnpem.br

The molecular mechanisms related to the bioconversion of aromatic compounds have been mainly studied in some model organisms, such as *Pseudomonas putida* KT2440[14,15] and *Sphingobium* sp. SYK-6[16,17], besides other species from the *Rhodococcus*[18] and *Burkholderia*[19,20] genera. These models have been isolated either from soil or industrial wastewater, so our understanding of how microorganisms from other ecological niches metabolize lignin-related compounds remains elusive. Moreover, how these molecules impact microbial behavior and physiology is still partially understood[21,22].

Some plant pathogens, such as *Xanthomonas* species, have a vast arsenal of enzymes to degrade components of the plant cell wall such as xyloglucan[23] and xylan[24], using the released carbohydrates as a source of carbon, energy, and stimuli[25]. However, little is known about their capacity and molecular strategies to metabolize other compounds available in the plant cell wall, especially the phenolics related to lignin, a major plant cell wall component[26–29]. One of the plant defense mechanisms against *Xanthomonas* infection is to increase the lignification of the plant cell wall, which implies an increased secretion of monolignols in the infection site[30]. Therefore, monolignol degradation could be an effective way for the pathogen to inhibit lignification. However, it is still unknown whether *Xanthomonas* species are biochemically capable of adopting such a strategy.

In this work, by combining RNA sequencing (RNA-seq) analysis, biochemical characterization, and gene knockout studies, we investigated the metabolism of lignin-related aromatics in the model phytopathogen *Xanthomonas citri subsp. citri* 306 (*X. citri* 306). Our data revealed missing steps and complete pathways for the catabolism of the three main lignin precursors, as well as reductive metabolic pathways and efflux approaches to cope with aryl aldehyde toxicity. This study also showed that lignin-related compounds activate transcriptional responses related to chemotaxis and flagellar-dependent motility in the phytopathogen, which might have an important role during the infection of the plant host. In summary, this work provides insights into the molecular mechanisms involved in plant-pathogen interactions and adds missing pieces to the known spectrum of molecular strategies for the bioconversion of lignin-related compounds.

Results

The model plant pathogen metabolizes a diverse range of lignin-related aromatics

To investigate if *X. citri* 306 can grow using lignin-related aromatic compounds as the main carbon source, we performed growth assays in minimal medium supplemented with 21 aromatic monomers, representative of H, G and S units (Fig. 1). Additionally, we analyzed two complex samples of lignin-derived compounds (LDC-I and LDC-II) produced from sugarcane bagasse (Supplementary Fig. 1 and Supplementary Table 1). The main aromatic monomers detected in the LDC-I sample were *p*-coumarate, ferulate, 4-hydroxybenzaldehyde, vanillin, along with organic acids (acetate, formate, lactate) and sugars (arabinose, glucose) (Supplementary Table 1). In LDC-II, the most abundant molecules detected were acetate and the aromatic compounds catechol, 3-methoxy-catechol, phenol, guaiacol and pyrogallol (Supplementary Table 1).

Under the tested conditions, only 4-hydroxybenzoate (4HBA), LDC-I and LDC-II supported the growth of *X. citri* 306 (Fig. 1, Fig. 2a and Supplementary Fig. 2). After 30 h of *X. citri* 306 growth in the LDC-I condition, we observed the total depletion of 4-hydroxybenzaldehyde, glucose, acetate, and formate (Supplementary Fig. 3). Lactate, succinate and arabinose were partially depleted (40–55%), with negligible depletion of hydroxycinnamic acids (Supplementary Fig. 3). In the LDC-II condition added by glucose (5 mmol L^{-1}), we observed the total depletion of hydroquinone and acetate, and partial depletion of 3-methoxycatechol (59%), 4-methylcatechol (40%), and catechol (11%) (Supplementary Fig. 3).

To evaluate whether the lack of growth in some conditions was due to toxicity or to insufficient carbon and energy supply, we repeated the assay supplementing the medium with glucose. Glucose supplementation allowed the growth of *X. citri* 306 in the presence of the model aromatic compounds, except for those displaying severe toxicity at 5 mmol L^{-1}, such as some aryl aldehydes (Fig. 1 and Supplementary Fig. 2). Overall, the presence of 5 mmol L^{-1} of the individual aromatic compounds decreased the growth rate when compared to the medium containing only glucose (XVM2m(G)), indicating toxicity. The aryl aldehydes displayed a more toxic effect than the correspondent aryl alcohols and aryl acids (Fig. 1 and Supplementary Fig. 2).

Next, we investigated if the aromatic compounds are depleted from the XVM2m(G) medium by *X. citri* 306, indicating they are either modified into other metabolites or funneled to the central carbon metabolism. For this purpose, we analyzed by HPLC the medium supernatant before and after bacterial growth using two input concentrations (millimolar and/or micromolar) of aromatic compounds (Fig. 1 and Supplementary Table 2). In micromolar aryl alcohol cultivations, only the three monolignols (*p*-coumaryl, coniferyl, and sinapyl alcohol) were effectively depleted from the medium by *X. citri* 306, generating oxidized metabolites detected in the medium (Fig. 1, Fig. 2b, and Supplementary Table 2). For all the tested aldehydes, a depletion higher than 70% was observed at the millimolar condition (except for 4-hydroxybenzaldehyde, <30%) (Fig. 1 and Supplementary Table 2). Among the aryl acids tested, *X. citri* 306 substantially depleted only 4-hydroxybenzoate. Together, these results indicate that *X. citri* 306 has pathways to metabolize aromatic compounds as complex as the monolignols and effectively metabolizes a more diverse range of aryl aldehydes compared to aryl alcohols and aryl acids provided in the culture medium (Fig. 1 and Supplementary Table 2).

Lignin-related compounds induce chemotaxis and flagellar-dependent motility

To investigate how *X. citri* 306 responds to lignin-related compounds and identify the metabolic pathways involved in their metabolism, we performed RNA-seq studies. Six model compounds (coniferyl alcohol, 4-hydroxybenzoate, 4-hydroxybenzaldehyde, vanillin, syringaldehyde, and benzaldehyde) were selected for the RNA-seq analyses as representatives of both different molecular structures (H, G and S-type units) and different entry points in the metabolic pathways. We also included in the tested conditions three complex mixtures of aromatic compounds (LDC-I, LDC-II, and aldehydes mix) (Supplementary Table 3).

A total of 278 to 1285 differentially expressed genes (DEGs) in the aromatic-containing conditions compared to the control XVM2m(G) were identified (Supplementary Data 1), evidencing the importance of these compounds in modulating various physiological processes. *X. citri* 306 discerned subtle structural variations among aromatic compounds, as shown by the incomplete overlap of upregulated genes in each condition (Fig. 3). For example, 4-hydroxybenzaldehyde activated the expression of around 400 genes that were not activated by 4-hydroxybenzoate, although they share a subset of about 200 genes activated by both (Fig. 3a). This discrepancy suggests that, although very similar in terms of molecular structure, 4-hydroxybenzaldehyde and 4-hydroxybenzoate can elicit substantially distinct transcriptional responses in *X. citri* 306, which might be related to the high toxicity of the aldehyde. Partially overlapping responses were also observed when comparing the same type of phenolic compound with distinct degrees of methoxylation (Fig. 3b) and complex mixtures with different compositions (Fig. 3c, Supplementary Table 1).

Although *X. citri* 306 recognizes subtle changes in the structure of lignin-related aromatic compounds, Gene Ontology (GO) enrichment analysis revealed that biological processes such as signal transduction,

Conditions	XVM2m	XVM2m(G)		Depletion (HPLC)	
	OD$_{600max}$	OD$_{600max}$	μ (h^{-1})	high conc.	low conc.
Aromatic-free control	0.03 ± 0.00	0.12 ± 0.01	0.44		N.A.
Aryl alcohols (mmol L^{-1})	5	5	5	2	0.05
p-coumaryl	0.00	0.05 ± 0.00	0.08		
coniferyl	0.00	0.11 ± 0.01	0.27		
sinapyl	0.00	toxic	toxic	N.A.	
4-hydroxybenzyl	0.00	0.13 ± 0.01	0.20		
vanillyl	0.00	0.10 ± 0.02	0.18		
syringyl	0.00	0.10 ± 0.00	0.12		
benzyl	0.00	0.11 ± 0.01	0.09		
Aryl aldehydes (mmol L^{-1})	5	5	5	1	0.05
p-coumaraldehyde	0.00	toxic	toxic		N.A.
coniferaldehyde	0.00	toxic	toxic		N.A.
sinapaldehyde	0.00	toxic	toxic		N.A.
4-hydroxybenzaldehyde	0.00	toxic	toxic		N.A.
vanillin	0.00	0.09 ± 0.00	0.08		N.A.
syringaldehyde	0.00	0.09 ± 0.00	0.07		N.A.
benzaldehyde	0.03 ± 0.00	0.11 ± 0.01	0.10		N.A.
Aryl acids (mmol L^{-1})	5	5	5	5	0.05
p-coumarate	0.00	0.11 ± 0.00	0.17		
ferulate	0.02 ± 0.00	0.07 ± 0.00	0.14		
sinapate	0.02 ± 0.00	0.05 ± 0.00	0.17		
4-hydroxybenzoate	0.11 ± 0.00	N.A.	N.A.		
vanillate	0.00	0.08 ± 0.02	0.13		
syringate	0.02 ± 0.00	0.11 ± 0.02	0.25		
benzoate	0.00	0.10 ± 0.02	0.15		
Lignin-derived complex mixtures (g L^{-1})					
LDC-I (1.0)	0.10 ± 0.00	N.A.	N.A.	N.A.	N.A.
LDC-II (0.3)	0.06 ± 0.00	N.A.	N.A.	N.A.	N.A.

Color bar legend: 0.01 — 0.5 — 1

Fig. 1 | Summary of growth conditions and analyses of aromatics toxicity and depletion. Growth assays in minimal medium (XVM2m) and XVM2m plus 5 mmol L^{-1} glucose (XVM2m(G)) supplemented with 5 mmol L^{-1} of lignin-related aromatic compounds or 1 g L^{-1} LDC-I or 0.3 g L^{-1} LDC-II. LDC = lignin-derived compounds. LDC-I is a residual liquid stream resulting from the acid precipitation of lignin from an alkaline liquor of sugarcane bagasse. LDC-II is a bio-oil rich in aromatic monomers, obtained by hydrothermal depolymerization of an alkaline lignin from sugarcane bagasse (Supplementary Fig. 1 and Supplementary Table 1). The red gradient represents differences in growth parameters from low (blank) to high (red). HPLC analyses were performed using XVM2m(G) medium plus 2 mmol L^{-1} aryl alcohols, 1 mmol L^{-1} aldehydes, 5 mmol L^{-1} acids, or 50 μmol L^{-1} of either aryl alcohols or acids. The blue color bar represents the percentage of aromatic compound depleted by *X. citri* 306 after around 15 h of growth (as detailed in Supplementary Table 2). Conc. concentration, N.A. not analyzed. Growth data are shown as mean ± SD of *n* = 3 or *n* = 4 biological replicates.

bacterial-type flagellum assembly, bacterial-type flagellum-dependent cell motility, and chemotaxis were enriched and upregulated in all conditions featuring lignin-related aromatics (Fig. 3d, Supplementary Fig. 4 and Data 2), except in the LDC-I and LDC-II conditions. This might be due to the low individual concentration of aromatic inducers or to the interference of non-aromatic molecules present in these mixtures (Supplementary Fig. 3).

In the conditions containing individual aromatic compounds, the upregulation of chemotaxis and flagellar genes, including *cheAZY* (*XAC1930-32*), *motAB* (*XAC3693-94*), *fliC* (*XAC1975*), *flgL* (*XAC1976*), *flgG* (*XAC1981*), and *flgE* (*XAC1983*) suggests that the sensing of lignin-related aromatics stimulates a motile state in *X. citri* 306 (Supplementary Data 1). This observation is consistent with the results of the co-expression analysis, where a co-expressed gene module (M1), mainly composed of genes involved in flagellar assembly and chemotaxis, was identified (Fig. 3e). Higher activity within this module became especially pronounced in the presence of 4-hydroxybenzaldehyde (4HBZ) and syringaldehyde (SYALD) while showing reduced activity under conditions involving complex lignin-derived samples (LDC-I and LDC-II), which can be due to the

comparatively lower concentration of individual aromatic compounds within these samples or to signal interference from other molecules present in these mixtures (Fig. 3e and Supplementary Fig. 3). Between the downregulated processes, translation was consistently enriched in all the tested conditions, except for 4HBA (Supplementary Fig. 4 and Data 2).

The first steps of monolignols catabolism are performed by aryl alcohol and aryl aldehyde dehydrogenases

The first step of coniferyl alcohol catabolism can be catalyzed by an NAD$^+$ dependent aryl alcohol dehydrogenase (ADH), generating coniferaldehyde, which is then converted to ferulate by a NAD$^+$-dependent aryl aldehyde dehydrogenase (ALDH)[32,33]. For p-coumaryl and sinapyl alcohols, this information is still missing, but considering their chemical similarity to coniferyl alcohol, we hypothesized that their catabolism might follow a similar pathway. Thus, to uncover the genes responsible for monolignols catabolism, we searched for ADH and ALDH genes upregulated in the presence of lignin-related compounds in *X. citri* 306. Based on their higher upregulation levels, genomic context, and the presence of common catalytic domains reported for
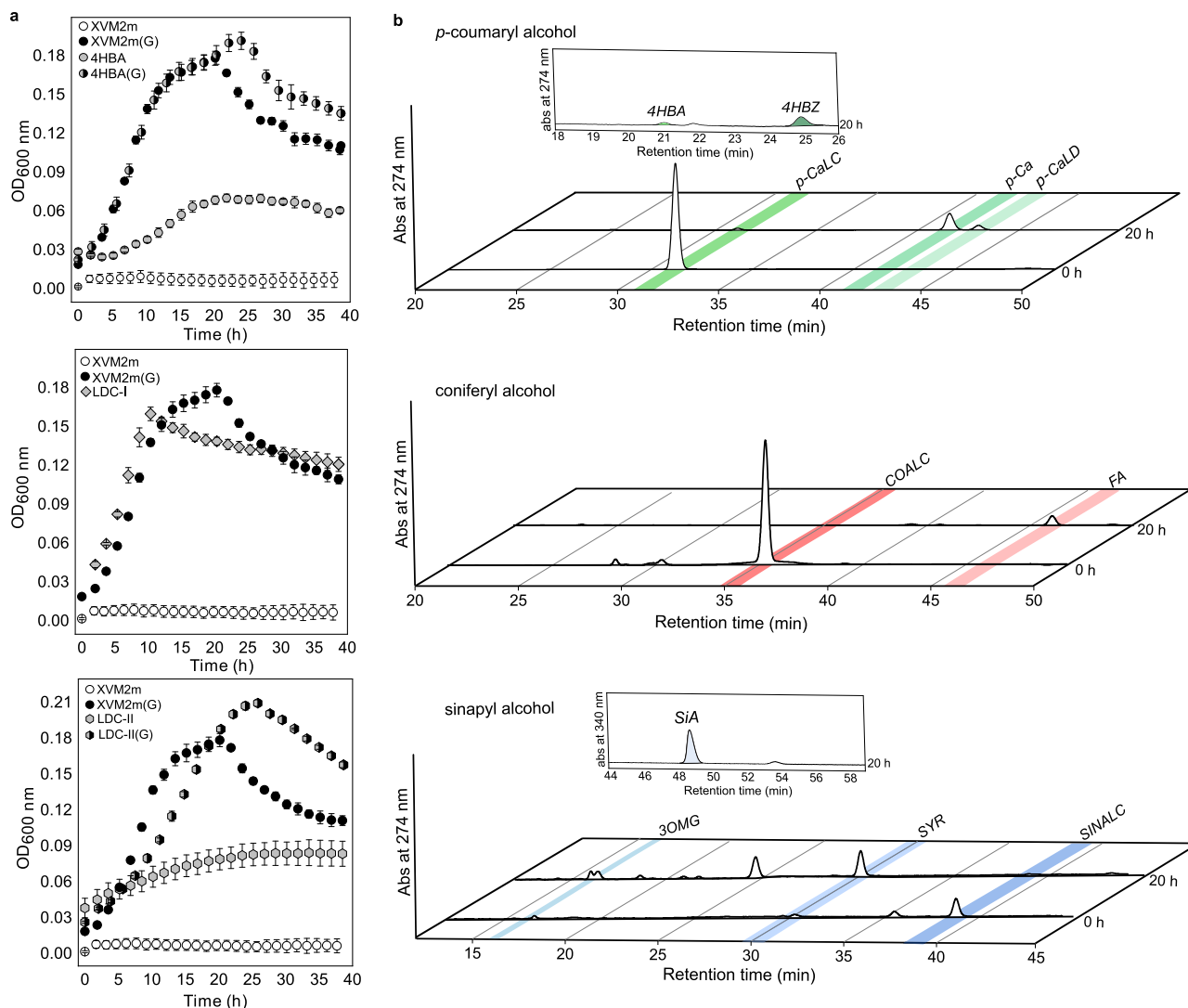
Fig. 2 | *X. citri* 306 grows using 4-hydroxybenzoate and lignin-derived compounds as carbon sources and metabolizes the three main monolignols. **a** Growth curves (hours) in XVM2m supplemented with 4-hydroxybenzoate (4HBA) and two complex mixtures of lignin-derived compounds (LDC-I and LDC-II). (G) − glucose-supplemented media. **b** HPLC chromatograms showing the depletion of the three monolignols and production of intermediate metabolites at 20 h post inoculation. 4HBZ 4-hydroxybenzaldehyde, *p*-CaLC *p*-coumaryl alcohol, *p*-Ca *p*-coumarate, *p*-CaLD *p*-coumaraldehyde, COALC coniferyl alcohol, FA ferulate, SiA sinapate, 3OMG 3-*O*-methylgallate, SYR syringate, SINALC sinapyl alcohol. Growth data are shown as mean ± SD of *n* = 3 biological replicates. Source data are provided as a source data file.

dehydrogenases active on aromatics, we selected eight ADH genes and three ALDH genes for cloning, heterologous expression, and biochemical activity screening (Fig. 4).

The activity screening revealed several alcohol and aldehyde dehydrogenases active on aromatic compounds, with variable preferences in terms of substrate, co-substrate (NAD(H)/NADP(H)) and reaction direction (oxidation and/or reduction), which will be detailed in later sections (Fig. 4 and Supplementary Tables 4, 5 and 6). Reaction products were confirmed by HPLC analyses, showing the aryl alcohol dehydrogenase activity of XAC0353, the aryl aldehyde dehydrogenase activity of XAC0129, XAC0354, and XAC0882 and the aryl aldehyde reductase activity of XAC1484 and XAC3477 (Supplementary Figs. 5–7 and Supplementary Table 6).

Between the screened genes, *XAC0353* and *XAC0354* called our attention because they are clustered in the genome and encode enzymatic activities compatible with the first steps of monolignols catabolism (Fig. 5a). *XAC0353* gene product displayed a NAD⁺-dependent alcohol dehydrogenase activity over the three main monolignols, besides cinnamyl and 4-hydroxybenzyl alcohols, whereas *XAC0354*

gene product showed a NAD⁺-dependent aldehyde dehydrogenase activity on H-, G- and S-type hydroxycinnamic aldehydes, suggesting a role in the second step of monolignols catabolism in *X. citri* 306 (Fig. 4).

To confirm these results, we measured the specific activity of XAC0353 and XAC0354 in several aromatic substrates (Fig. 5b, c). As expected, XAC0353 converted aryl alcohols in their respective aldehydes using NAD⁺ as co-substrate, showing higher specific activity over the monolignols compared to benzyl alcohol derivatives (Fig. 5b). Accordingly, the aldehyde dehydrogenase encoded by *XAC0354* transformed aryl aldehydes in their respective acids in a NAD⁺-dependent manner, showing the highest specific activity on coniferaldehyde (100%), followed by *p*-coumaraldehyde (51%) and sinapaldehyde (34%), compared to the other tested aromatics (Fig. 5c). Both enzymes were assayed in similar conditions and displayed similar specific activity rates over their best substrates (monolignols and the respective aldehydes), which might be advantageous to avoid the accumulation of the intermediate aldehyde metabolites during the serial action of these enzymes.
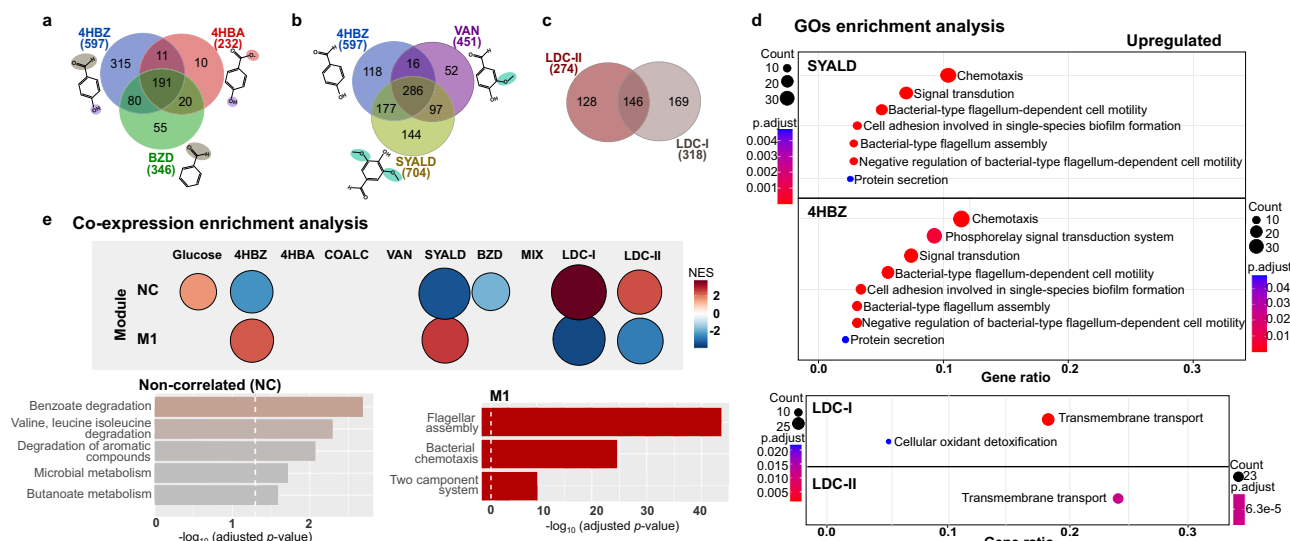
Fig. 3 | Transcriptional responses triggered by lignin-related aromatic compounds. Venn diagrams comparing the distribution of unique and shared upregulated genes in conditions containing (**a**) 4-hydroxybenzaldehyde (4HBZ), benzaldehyde (BZD), or 4-hydroxybenzoate (4HBA) (**b**) 4HBZ, vanillin (VAN), or syringaldehyde (SYALD) and (**c**) Lignin-derived compounds - LDC-I and LDC-II conditions. **d** Gene ontology (GO) enrichment analysis of up regulated genes based on one-side Fisher's exact test as implemented in the clusterProfiler R package[77]. Circles' size and color represent the counts and adjusted p-values, respectively. Gene ratio corresponds to the number of DEGs related to a GO term divided by the total number of genes associated with that GO term in the *X. citri* 306 genome. The differential expression in each condition was compared to XVM2m(G) following the criteria log$_2$ Fold Change ≥1 and p-adjusted ≤ 0.05. The analysis of other conditions is in Supplementary Fig. 4. **e** Gene set enrichment analysis based on weighted Kolmogorov–Smirnov statistic and Over Representation Analysis (ORA) to define modules function as implemented in the CEMiTool package[31]. The size and intensity of the circles correspond to the normalized enrichment score (NES) for the module in each condition, indicating biological functions enriched in each module. Positive NES reflects transcriptional activity above the median, whereas negative NES corresponds to transcriptional activity below the median in each condition. COALC coniferyl alcohol, MIX aldehyde mixture. Source data are provided as a source data file.

To validate the in vivo involvement of these enzymes in monolignol catabolism, we characterized individual knockout strains of *XAC0353* and *XAC0354* genes. Growth assays in XVM2m(G) containing 50 µmol L^{-1} monolignols revealed that *XAC0353* deletion impaired monolignols consumption and decreased the excretion of hydroxycinnamic acids, which are the second intermediate metabolites of monolignols catabolism (Fig. 5d). The excretion of the first intermediates, the aryl aldehydes, was not observed, indicating they are rapidly converted to the respective acids. Regarding sinapyl alcohol, although we did not detect its presence in the growth assays, likely due to instability issues[34], we noticed a lower amount of sinapate being excreted by the *KO53* strain compared to the WT, which is consistent with the results observed for the other two monolignols (Fig. 5d). Of note, the *KO53* strain still retained a partial capacity to convert monolignols up to hydroxycinnamic acids, indicating the existence of at least another gene encoding for an enzyme with aryl alcohol dehydrogenase activity. Together, these results show that the deletion of *XAC0353* gene partially compromises the catabolism of the monolignols, leading to a lower production of hydroxycinnamic acids, whose conversion to the next metabolite might be a limiting factor in *X. citri* 306.

The growth assays using 50 µmol L^{-1} of hydroxycinnamic aldehydes showed no detectable excretion of hydroxycinnamic acids by the *KO54* strain (Δ*XAC0354*), supporting a crucial role for this gene in the conversion of these aldehydes into acids (Fig. 5e). On the other hand, a higher amount of excreted monolignols (*p*-coumaryl and coniferyl alcohol) was observed for the *KO54* strain compared to the WT, indicating that the deletion of the *XAC0354* gene triggered a metabolic shift toward reactions that reduce the aldehydes into their respective monolignols. This is consistent with the complete depletion of the hydroxycinnamic aldehydes by the WT and the *KO54* strain, although by preferentially following opposite metabolic directions.

Syringaldehyde induces the catabolism of hydroxycinnamic acids
Although *X. citri* 306 did not utilize hydroxycinnamic acids (HCAs) when available in the growth medium (Supplementary Table 2), HPLC analysis indicated that they are produced as intermediate metabolites of monolignols catabolism, along with smaller intermediates resultant from their degradation such as 4-hydroxybenzoate and syringate (Fig. 2b). Thus, these findings suggest that *X. citri* 306 possesses the necessary enzymatic systems for metabolizing HCAs produced intracellularly.

Genome mining revealed a gene cluster (*XAC0881-84*) homologous to the *hca* gene cluster responsible for the degradation of *p*-coumarate, ferulate, and sinapate in *Xanthomonas campestris* pv. *campestris*[29] (Supplementary Fig. 8a). In our RNA-seq data, this cluster was upregulated in the presence of syringaldehyde, an intermediate of the sinapate catabolism (Supplementary Fig. 8b and Data 1). *E. coli* cells co-expressing *XAC0881* and *XAC0883* transformed *p*-coumarate, ferulate and sinapate respectively into 4-hydroxybenzaldehyde, vanillin and syringaldehyde, experimentally demonstrating the predicted activities of these genes (Fig. 6).

The *hca* gene cluster also harbors *XAC0882*, which encodes for an aldehyde dehydrogenase active on 4-hydroxybenzaldehyde, vanillin, and syringaldehyde, according to our activity screening results and HPLC data (Fig. 4, Fig. 6b–d and Supplementary Table 6). This activity is compatible with the third step of HCAs catabolism, corroborating the involvement of the *XAC0881-84* cluster in the metabolism of HCAs in *X. citri* 306. Besides *XAC0882*, another gene (*XAC0129*) also encodes for an aldehyde dehydrogenase active on aldehydes derived from HCAs catabolism (Fig. 4 and Supplementary Table 6), implying that there is a functional redundancy for this metabolic step in *X. citri* 306.

Novel reductive pathways for aryl aldehydes detoxification
Among the putative ADHs screened in the previous section, XAC1484 and XAC3477 displayed a NADPH-dependent reductase
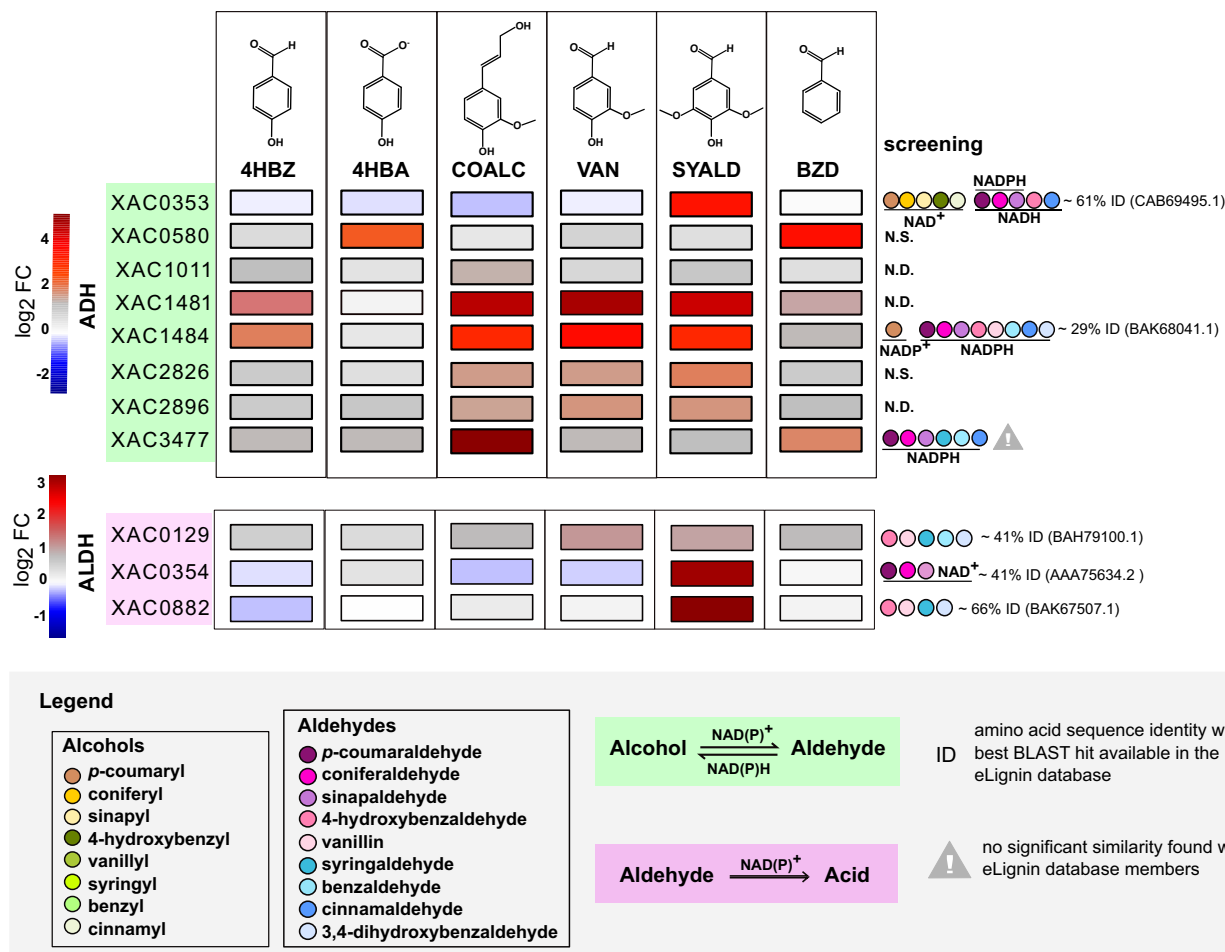
Fig. 4 | Transcriptomic analysis and activity screening reveal novel ADH and ALDH enzymes active on aromatic compounds. The heatmap presents transcription levels (log$_2$FC = log$_2$ Fold Change), comparing each growth condition to the reference (XVM2M(G)) from at least n = 3 biological replicates. Genes were classified as upregulated according to the following criteria: log$_2$FC ≥ 1, with an p-adjusted ≤ 0.05. log$_2$ Fold Change was calculated using edgeR[76] based on likelihood ratio test within a negative binomial generalized log-linear model framework. The activity screening was performed using purified enzymes, as detailed in Supplementary Data 3 and Supplementary Tables 4 and 5, or whole cells assays (XAC0129 and XAC0882) as detailed in Supplementary Table 6. N.S. indicates proteins that were insoluble in *E. coli*. N.D. indicates soluble proteins that did not display activity in the tested conditions. % ID = amino acid sequence identity with the most similar enzyme sequence listed in the eLignin database (GenBank accession number in parentheses). 4HBZ 4-hydroxybenzaldehyde, 4HBA 4-hydroxybenzoate, COALC coniferyl alcohol, VAN vanillin, SYALD syringaldehyde, BZD benzaldehyde. The ADH (green box) enzymes were subjected to screening for both direct and reverse reactions, using the corresponding substrates listed in the legend. The ALDH (pink box) enzymes were screened only for aldehyde dehydrogenation. Colored circles indicate the substrates and labels indicate the respective co-substrate with which the enzymes were active.

activity, catalyzing the reduction of aryl aldehydes in their respective alcohols, being active on derivatives of both cinnamyl and benzyl aldehydes (Fig. 4, Supplementary Table 4 and Supplementary Fig. 6 and 7). Curiously, XAC0353 also displayed a similar activity, but in a NAD(P)H-dependent manner, indicating that it might also contribute to the reduction of aryl aldehydes depending on the intracellular balance of NAD$^+$/NAD(P)H and aryl alcohol/aryl aldehyde substrates. Indeed, *X. citri* 306 growing on a mixture of aldehydes at millimolar level showed to be capable of converting most of them into aryl alcohols, which were then exported to the extracellular medium (Fig. 7a). In such condition, the genes *XAC1484* and *XAC3477* were upregulated (Supplementary Data 1). This transcriptional evidence, along with biochemical activity data (Fig. 4, Supplementary Table 4 and Supplementary Fig. 6 and 7), support that XAC1484 and XAC3477 play a role in converting aryl aldehydes into their respective aryl alcohols in vivo. Moreover, *XAC1484* is clustered with the genes *XAC1482-83-85*, which encodes for a putative Resistance-Nodulation-Division (RND) multidrug efflux transporter, which might be involved in the excretion of the aryl alcohols produced by XAC1484.

Since two metabolic directions, reductive or oxidative, are possible to the bioconversion of aryl aldehydes in *X. citri* 306, we argued how the concentration of these compounds affects their metabolic fate. To address this issue, we grew *X. citri* 306 in minimal medium supplemented with increasing concentrations of coniferaldehyde, used as a representative compound, and quantified its depletion as well as the production of intermediate metabolites excreted after 15 h of growth (Fig. 7b). In the lowest coniferaldehyde concentration tested, only ferulate excretion was detected. Increasing the concentration of coniferaldehyde, coniferyl alcohol became the predominant excreted metabolite (Fig. 7b). Thus, these results indicate that increasing concentrations of aryl aldehydes activate reductive pathways, providing an additional mechanism to cope with toxicity, which was generally higher for aryl aldehydes than to their alcohol or acid counterparts (Fig. 1).

H and G-type compounds are metabolic funneled via protocatechuate ortho-cleavage
The final step in the funneling pathways of H- and G-type monomers forms protocatechuate (PCA) (Fig. 8a)[7,10]. In the case of H-type compounds, it is known that 4-hydroxybenzoate (4HBA) is converted to
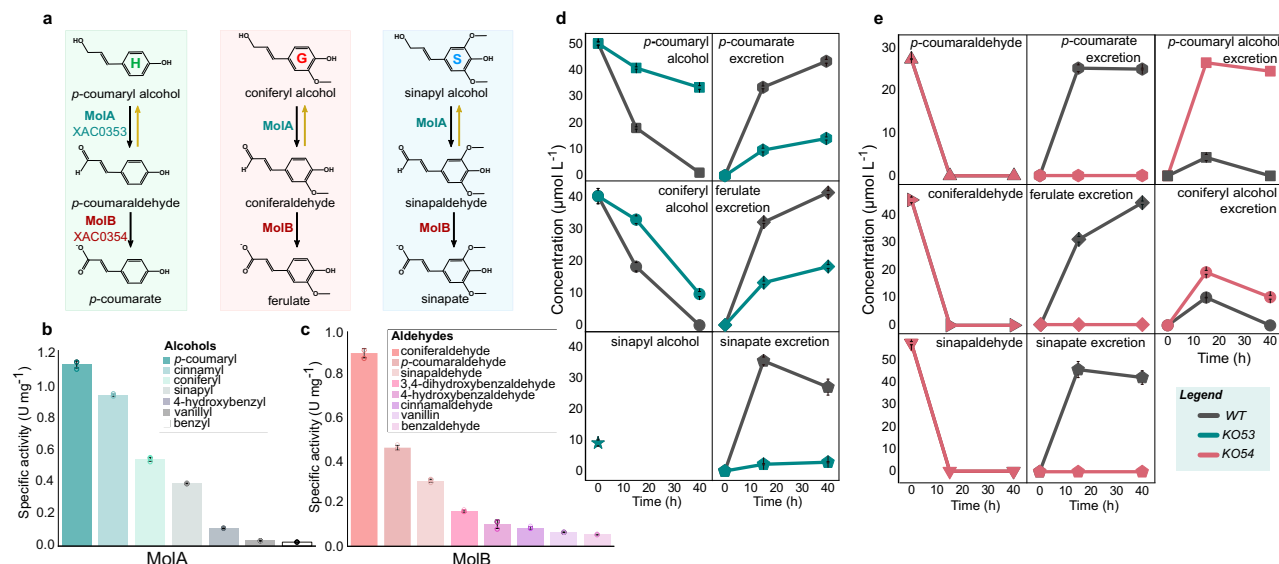
Fig. 5 | XAC0353 and XAC0354 play a role in the first steps of monolignols catabolism. a Schematic of the proposed *p*-coumaryl alcohol (green H), coniferyl alcohol (red G), and sinapyl alcohol (blue S) bioconversion pathway in *X. citri* 306. Yellow arrows indicate the reaction of aldehydes reduction (reductive pathways), and black arrows indicate oxidative steps. **b, c** Specific activity for the dehydrogenation of aryl alcohols and aryl aldehydes catalyzed by the enzymes MolA and MolB encoded by *XAC0353* and *XAC0354*, respectively. The activity was calculated based on NADH production, quantified by HPLC. U μmol min⁻¹. **d** HPLC analysis of the consumption of H, G, and S monolignols and the excretion of intermediate metabolites (hydroxycinnamic acids) by the *WT* (dark gray) and *KO53* knockout (green) strains. Sinapyl alcohol was not detected during bacterial growth, likely due to instability issues (star). **e** HPLC analysis of the consumption of H, G, and S hydroxycinnamic aldehydes and the excretion of metabolites (hydroxycinnamic acids or monolignols) by the *WT* (dark gray) and *KO54* knockout (pink) strains. Sinapyl alcohol excretion was not detected, which we attribute to its instability as previously reported[34]. In all panels, data are shown as mean ± SD of *n* = 3 biological replicates. Source data are provided as a source data file.

PCA by the enzyme *p*-hydroxybenzoate hydroxylase (EC 1.14.13.2)[35]. In *X. citri* 306 genome, we found a gene (*XAC0356*) encoding for a protein homologous to the *p*-hydroxybenzoate hydroxylase (PobA) from *Pseudomonas aeruginosa*[36] and from *X. campestris*[28]. *XAC0356* was upregulated in the 4HBA condition (Fig. 8b). Its deletion abolished *X. citri* 306 growth in a medium containing 4HBA as the primary carbon source. It also disrupted 4HBA consumption when the Δ*pobA* strain grew in a medium containing glucose and 4HBA (Fig. 8c, d). These results demonstrate that *XAC0356* encodes a functional PobA and is essential for the metabolism of 4HBA in *X. citri* 306.

In the last step of the funneling pathway for G-type monomers, vanillate needs to be demethylated to form PCA[7,12]. Three main types of *O*-demethylase systems have been reported in the literature so far: tetrahydrofolate (THF)-dependent enzymes, Rieske-type oxygenases and cytochromes P450 oxygenases[37,38]. In the *X. citri* 306 genome, we found genes homologous to Rieske-type oxygenases (*XAC0311* and *XAC0363*) clustered with genes encoding for putative reductases (*XAC0310* and *XAC0362*), as well as a predicted P450 gene (*XAC3170*, <24% sequence identity with GcoA)[39]. In the vanillin condition, only *XAC0362-63* were upregulated (Fig. 8e). The deletion of this gene pair (but not of *XAC0310-11*) resulted in vanillate accumulation after 40 h of bacterial cultivation in a medium containing vanillin (Fig. 8f). Enzymatic assays showed that XAC0362-63 display vanillate-*O*-demethylase activity, further supporting a role for these proteins in converting vanillate into PCA (Supplementary Fig. 9).

Once formed, PCA can be cleaved in three different positions: 2,3-cleavage, 4,5-(meta-cleavage), or 3,4-(ortho-cleavage)[7]. Conditions containing H and G-type aromatic compounds induced the expression of a gene cluster (*XAC0364-71*) homologous to the *pcaIJFHGBDC* cluster previously characterized in *X. campestris*[28], suggesting that *X. citri* 306 conserves the same PCA 3,4-(ortho-cleavage) pathway (Fig. 8g and Supplementary Data 1).

Benzaldehyde metabolism generates dead-end products

In the RNA-seq data, we observed that benzaldehyde upregulates the gene *pobA* and the *pcaIJFHGBDC* cluster (Fig. 8g and Supplementary Data 1) and this intriguing result prompted us to investigate if its metabolism involves these genes. To test this hypothesis, we performed activity assays and gene knockout studies targeting PobA (hydroxylase) and PcaHG (the first enzyme of the PCA ortho-cleavage pathway). PobA showed no detectable activity towards benzoate (Supplementary Fig. 10a). The Δ*pobA* strain was similar to the WT strain in the conversion of benzaldehyde and excretion of its respective intermediate metabolites (Supplementary Fig. 10b). The Δ*pcaHG* strain accumulated the intermediate metabolite PCA in positive controls containing 4-hydroxybenzoate or vanillin, but the same was not observed in presence of benzaldehyde (Supplementary Fig. 10c–e). Together, these results indicate that benzaldehyde metabolism has no connection with PobA and the PCA ortho-cleavage pathway, despite its role as an inducer of *pobA* and *pca* genes expression. Since benzaldehyde was depleted from the medium by *X. citri* 306 (Fig. 1), and it is a substrate of aryl aldehyde reductases (XAC1484 and XAC3477) (Supplementary Table 4) and dehydrogenase (XAC0129) (Supplementary Table 6), we then hypothesized that benzaldehyde metabolism generates benzyl alcohol and benzoate as dead-end products, which was confirmed by quantitative analysis using HPLC (Supplementary Fig. 10f).

Transcriptome response to syringaldehyde reveals the presence of a complete pathway for S-type lignin monomers catabolism

In nature, the catabolism of S-type monomers can follow at least three different pathways, categorized according to the generated intermediates: gallate (I), 2-pyrone-4,6-dicarboxylate (PDC) (II), or 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (III). In *X. citri* 306, we found genes (*XAC0882, XAC0878, XAC4155, XAC4156,* and *XAC4157*) that were upregulated in the presence of syringaldehyde (Fig. 9a) and are homologous to *desV, desB, ligK, ligU* and *ligJ* involved in the
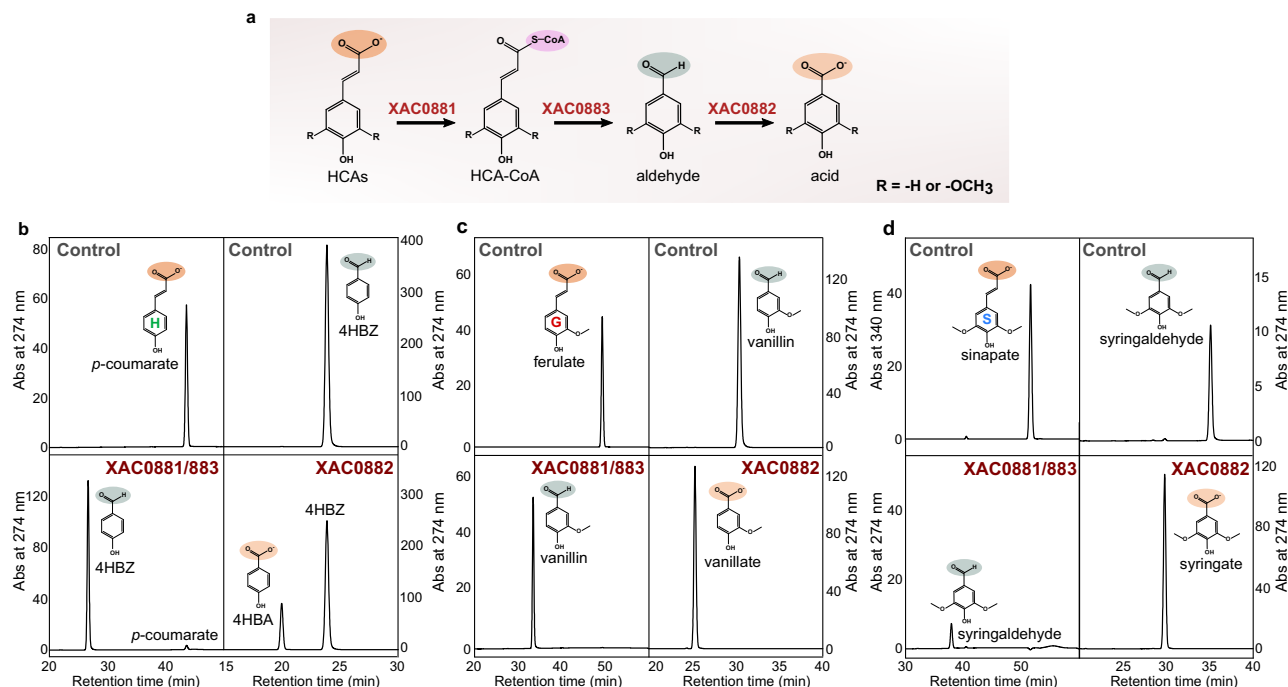
Fig. 6 | Hydroxycinnamic acids catabolism is encoded by the *hca* gene cluster. **a** Schematic representation of the degradation pathways proposed for HCAs in *X. citri* 306. The chemical group modified after each reaction (arrow) is highlighted with different colors. **b**–**d** Representative HPLC chromatograms of control reactions with *E. coli* cells transformed with empty vectors, showing only the substrates (top), and reactions with *E. coli* cells co-expressing XAC0881 and XAC0883 genes or expressing only XAC0882, showing the detected products (bottom). The peaks were assigned to each molecule based on comparison with analytical standards. Chemical groups are highlighted following the same color code of panel **a**. The whole-cell assays were performed with *n* = 2 biological replicates, as detailed in Supplementary Table 6 and 7. Source data are provided as a source data file.

catabolism of syringaldehyde via gallate in *Sphingobium* sp. SYK-6[12,40] (Fig. 9b and Supplementary Table 8).

The formation of gallate as an intermediate relies on a two-step *O*-demethylation process, catalyzed by tetrahydrofolate-dependent *O*-demethylases in *Sphingobium* sp. SYK-6[41]. However, no homologous to tetrahydrofolate-dependent *O*-demethylases was detected within the genome of *X. citri* 306. Conversely, the two gene pairs *XAC0362-63* and *XAC0310-11*, homologous to Rieske-type *O*-demethylases, were upregulated in presence of syringaldehyde (Fig. 9a), indicating a possible role in the *O*-demethylation of syringate (SYR) and 3-O-methylgallate (3OMG).

Growth assays with the knockout strains of genes *XAC0310−11* (*KO310-11*) and *XAC0362-63* (*KO362-63*) in a minimal medium containing syringaldehyde revealed a higher accumulation of syringate by the *KO362-63* strain, but not by *KO310-11* (Fig. 9c). This observation suggests a role for XAC0362-63 in syringate demethylation. Whole-cell activity tests corroborated this finding, showing the production of 3OMG from syringate by *E. coli* cells co-expressing *XAC0362-63*. They also revealed that XAC0362-63 converts 3OMG into gallate (Fig. 9d, e).

The remaining pathway leading gallate up to the TCA cycle was inferred based on genome mining analysis, homology inference with previously characterized genes and syringaldehyde-specific transcriptional activation (Fig. 9a, b). It encompasses the putative gallate dioxygenase encoded by the genes *XAC0878* (*ligB, β-chain*) and *XAC0879* (*ligA, α-chain*), which are homologs to the respective N-terminal and C-terminal domains of gallate dioxygenases from *P. putida* KT2440 (GalA)[42] and *Sphingobium* SYK-6 (DesB)[43] (Fig. 9b and Supplementary Fig. 11). Next steps are probably encoded by the *XAC4155-56-57* genes, which are homologous to the *ligK-ligU-ligJ* genes previously characterized in *Sphingobium* sp. SYK-6[12] (Fig. 9a, b).

Discussion
This study uncovers complete pathways for the catabolism of the three main lignin precursors in the model plant pathogen *X. citri* 306

(Fig. 10). So far, only the microbial catabolism of coniferyl alcohol has been reported[7], but in the context of a pathway having eugenol as the first substrate[44]. Our study demonstrates the existence of catabolic pathways starting from coniferyl alcohol as well as *p*-coumaryl and sinapyl alcohols, adding new pieces in the puzzle of microbial metabolic pathways for lignin-related aromatic compounds.

According to our data, the catabolism of monolignols in *X. citri* 306 starts with their uptake, probably facilitated by the putative outer membrane transporter MolK (*XAC0352*). MolK belongs to family COG4313, which has been implicated in the uptake of hydrophobic molecules[45]. Next, the monolignols are dehydrogenated to aldehydes and subsequently to hydroxycinnamic acids (HCAs) mainly by two NAD+-dependent enzymes: the aryl alcohol dehydrogenase MolA (*XAC0353*) followed by the aryl aldehyde dehydrogenase MolB (*XAC0354*) (Fig. 10).

MolA is closely related to the CalA enzyme (61% sequence identity, 99% query cover) from *Pseudomonas* sp. HR199[46]. CalA has been proposed to compose a pathway dedicated to the catabolism of eugenol in *Pseudomonas* sp. HR199, along with CalB, a coniferaldehyde dehydrogenase distantly related to MolB (30% sequence identity, 70% query cover)[32]. MolA and MolB co-occur in several plant-pathogenic, plant-symbiotic and environmental bacteria from the Xanthomonadales, Pseudomonadales, Burkholderiales and Rhizobiales orders, indicating that their biological roles go beyond the plant-pathogen context (Supplementary Fig. 12). Homologs closer to CalB were found exclusively in some *Pseudomonas* species and the triple co-occurrence of *molA/calA*, *molB* and *calB* genes was observed in some Pseudomonadales genomes, remaining to be determined its biological meaning.

Back to the pathways found in *X. citri*, after the action of MolA and MolB enzymes, the HCAs produced may be either excreted or converted into their respective hydroxybenzyl aldehyde derivatives via the CoA-dependent non-β-oxidation pathway encoded by the gene cluster (*XAC0881-83*) (Fig. 10). In contrast to *X. campestris*, which uptakes and
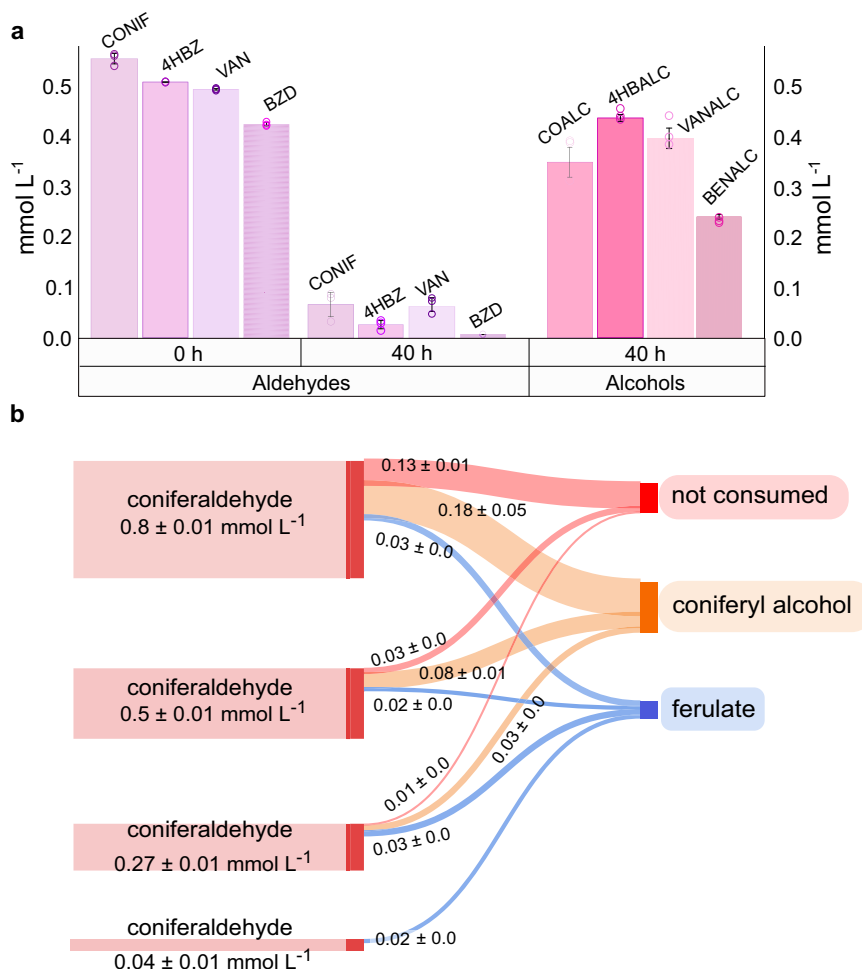
Fig. 7 | Reductive approaches provide an alternative route for aryl aldehydes detoxification. a HPLC analysis showing the consumption of a mixture of aldehydes (CONIF, 4HBZ, VAN, and BZD) and their respective alcohols excreted to the extracellular environment and detected after 40 h post inoculation. **b** Sankey graph (https://sankeymatic.com) of HPLC quantification of coniferaldehyde consumption and excreted metabolites after 15 h post inoculation. CONIF coniferaldehyde, BZD benzaldehyde, COALC coniferyl alcohol, VAN vanillin, 4HBZ 4-hydroxybenzaldehyde, VANALC vanillyl alcohol, 4HBALC 4-hydroxybenzyl alcohol, BENALC benzyl alcohol. Data are shown as mean ± SD of $n = 3$ biological replicates. Source data are provided as a source data file.

metabolizes HCAs[29], *X. citri* 306 apparently metabolizes only HCAs produced intracellularly, likely due to the lack of transporters for their uptake. This adaptation correlates with studies showing that HCAs are commonly found in citrus fruits in their conjugated forms (ester- or glycoside-bond), with minimal concentrations in their free form[47,48]. Part of the HCAs produced intracellularly is excreted by the cell, implying that HCAs deacetylation is probably a metabolic bottleneck to the flux towards the central carbon metabolism in *X. citri* 306.

Next, the H-G-S hydroxybenzyl aldehydes are converted to hydroxybenzoic acids by aryl aldehyde dehydrogenases, including XAC0882 and XAC0129. Then, 4HBA (H-type subunit) undergoes hydroxylation by the PobA enzyme (*XAC0356*) while vanillate (G-subunit) is *O*-demethylated by a Riske-type-oxygenase-reductase system VanAB (*XAC0363-62*), both steps resulting in the formation of PCA. Then, the PCA ring is ortho-cleaved and converted into β-ketoadipate by PcaHGBCD enzymes (*XAC0367-68-69-71-70*). Finally, PcaIJF enzymes (*XAC0364-65-66*) complete the conversions steps towards the tricarboxylic acid cycle (Fig. 10).

On another branch, syringate (S-type subunit) is *O*-demethylated to 3OMG and then to gallate by VanAB (*XAC0363-XAC0362*). Gallate is likely converted to 4-oxalomesaconate (OMA) by the action of LigAB (*XAC0878-79*) and follows a pathway up to pyruvate and oxaloacetate by enzymes encoded by the *XAC4155-56-57* genes, which are homologous to *ligK-ligU-ligJ* from *Sphingobium* sp. SYK-6[12] (Figs. 9b and 10).

This pathway was predicted based on transcriptional data and homology inference so future studies will be required to confirm the pathway and the metabolite intermediates leading gallate up to the TCA cycle in *Xanthomonas*.

Besides monolignols, *X. citri* 306 also uptakes their hydroxycinnamic and hydroxybenzoic aldehyde derivatives, driving them to two possible metabolic fates: oxidation to acids or reduction to alcohols. Two NADPH-dependent enzymes (XAC1484 and XAC3477) contribute to converting aryl aldehydes into aryl alcohols and the putative XAC1482-83-85 efflux system probably facilitates the aryl alcohol excretion. XAC1484 belongs to the Short-chain Dehydrogenases/ Reductases (SDR) family whereas XAC3477 belongs to the Aldo/Keto Reductase (AKR) family, contrasting to the Medium-chain Dehydrogenase/Reductase (MDR) family members reported to play a role in aryl aldehyde reduction in other bacterial species[49–52]. XAC1484 and XAC3477 share less than 41% sequence identity to functionally characterized enzymes, according to BLAST searches at the Swiss-Prot database[53]. Thus, they represent novel reductases with potential biotechnological applications in aryl aldehydes detoxification[54] and production of value-added aryl alcohols[55–57].

The presence of the phenolic group seems to be mandatory for the complete catabolism of lignin-related compounds, since benzaldehyde, a non-phenolic aromatic, was converted to dead-end products (benzyl alcohol and benzoate) (Fig. 10). Other adaptive responses
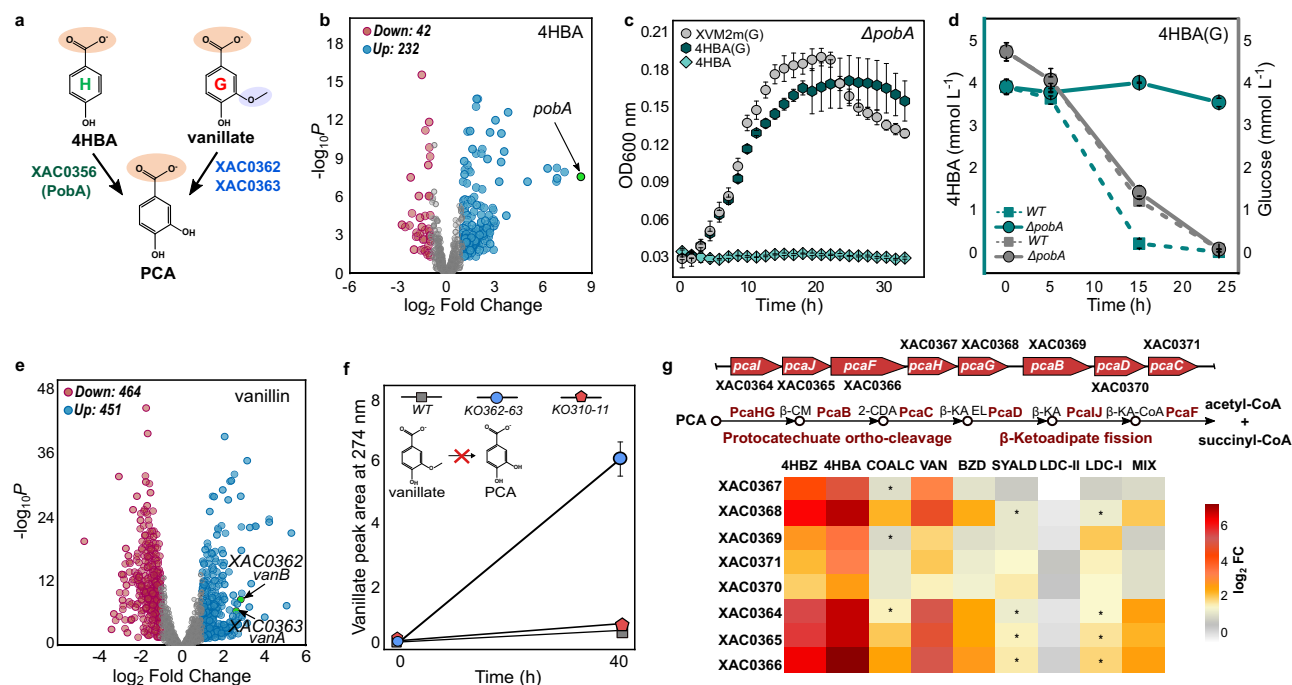
Fig. 8 | H and G-type monomers are funneled to the protocatechuate ortho-cleavage pathway. a Representative scheme of 4HBA and vanillate conversion into protocatechuate (PCA) in *X. citri* 306. **b** Volcano plot of RNA-seq data highlighting the *XAC0356* (*pobA*) gene upregulated in the 4HBA condition. Differential gene expression was calculated by edgeR[76] according to likelihood ratio test within a negative binomial generalized log-linear model framework. Genes with |log$_2$FC|>1 and adjusted *p*-value < 0.05 were assigned as DEGs. **c** Growth curve of *pobA* knockout strain on XVM2m supplemented with only glucose (G, light gray), only 4HBA (light green), or with HBA and glucose (4HBA(G), dark green). **d** 4HBA (green, left *y* axis) and glucose (gray, right *y* axis) consumption during the growth of WT (dashed lines) and Δ*pobA* strains (continuous lines) in the XVM2m medium supplemented with 4HBA and glucose. **e** Volcano plot of RNA-seq data highlighting the *XAC0362-63* genes upregulated in the vanillin condition. Statistics were calculated as described in panel **b**. **f** HPLC analysis showing vanillate accumulation only by the Δ*XAC0362-63* (*KO362-63*) strain. **g** Schematic representation of the genomic

context of *pca* genes, the corresponding metabolic steps, and a heat map of RNA-seq data (log$_2$ Fold Change) showing the *pca* genes up-regulated under conditions containing H and G-type aromatic compounds as well as benzaldehyde. log$_2$ Fold Change was calculated using edgeR[76] based on likelihood ratio test within a negative binomial generalized log-linear model framework. COALC coniferyl alcohol, VAN vanillin, 4HBA 4-hydroxybenzoate, 4HBZ 4-hydroxybenzaldehyde, BZD benzaldehyde (BZD), SYALD syringaldehyde, PCA protocatechuate, β-CM β-carboxy-*cis,cis*-muconate, 2-CDA γ-carboxymuconolactone, β-KA EL β-ketoadipate enol-lactone, β-KA β-ketoadipate, β-KA-CoA β-ketoadipyl-CoA. Genes were considered up-regulated according to the criteria, log$_2$ Fold Change ≥ 1, with adjusted *p*-value ≤ 0.05. (*) indicate genes with log$_2$ Fold Change ≥ 1 that do not fit the adjusted *p*-value ≤ 0.05 criterion. In **c**, **d** and **f** panels symbols and error bars represent mean ± SD of *n* = 3 biological replicates. Source data are provided as a source data file.

induced by aromatics include the upregulation of genes associated with chemotaxis and flagellar assembly, which might benefits *X. citri* fitness during host colonization. In *X. campestris*, the catabolism of plant-derived phenolic compounds has been shown to be important for virulence[27,28] and the induction of monolignols biosynthesis and plant cell wall lignification has been proved to be important host defense mechanisms[30,58,59]. Thus, the capacity of *X. citri* 306 to sequester and metabolize key precursors of lignin biosynthesis may locally disrupt plant defense mechanisms in the benefit of the bacterium. This feature seems to be shared with other plant pathogens, as suggested by the conservation of the *molRKAB* operon in other *Xanthomonas* phytopathogenic species and of the *molA* and *molB* genes in phytopathogens from other genera (Supplementary Fig. 13).

In short, the metabolism of aromatic compounds in *X. citri* 306 converges to the central carbon metabolism but displays escape routes (efflux transporters) and reductases that likely function as complementary strategies to rapidly detoxify aryl aldehydes. Excretion of aryl acids or aryl alcohols by *X. citri* might serve as both a detoxification strategy and a mechanism to further hamper lignin biosynthesis, as these excreted compounds possibly compete with lignin precursors and prevent their oxidation by the plant laccases. These molecular strategies might inspire metabolic engineering approaches for the redesign of lignin biosynthesis in plants, aiming to facilitate biomass saccharification without compromising plant health[60,61]. They

also may serve as hotspots for the development of new treatments against plant diseases. From the industrial point of view, the knowledge provided here may support the development of microbial strains more tolerant to aryl aldehydes or proficient on converting lignin-derived compounds from agro-industrial side-streams into valuable bioproducts.

Methods

Bacterial strains and culture conditions

X. citri 306 was grown at 30 °C, 200 rpm, in LBON medium (10 g L^{-1} bacto peptone and 5 g L^{-1} yeast extract), minimal medium XVM2m (20 mmol L^{-1} NaCl, 10 mmol L^{-1} (NH$_4$)$_2$SO$_4$, 1 mmol L^{-1} CaCl$_2$, 0.01 mmol L^{-1} FeSO$_4$.7 H$_2$O, 5 mmol L^{-1} MgSO$_4$, 0.16 mmol L^{-1} KH$_2$PO$_4$, 0.32 mmol L^{-1} K$_2$HPO$_4$, 0.03% m V^{-1} casamino acids, pH 6.7) supplemented with different carbon sources, as better described in the next sections. *Escherichia coli* DH5α™ was used for DNA cloning, and *E. coli* BL21(DE3)-ΔslyD-pRARE2 or *E. coli* SHuffle® T7 Express lysY (New England Biolabs) were used for heterologous protein expression. *E. coli* strains were grown at 37 °C (or 20 °C), 200 rpm in LB medium (10 g L^{-1} bacto peptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl), 2xYT (16 g L^{-1} peptone, 10 g L^{-1} yeast extract and 5 g L^{-1} NaCl), or in M9 minimal medium (6.78 g L^{-1} Na$_2$HPO$_4$, 3 g L^{-1} KH$_2$PO$_4$, 0.5 g L^{-1} NaCl, 1 g L^{-1} NH$_4$Cl, 2 mmol L^{-1} MgSO$_4$, 100 μmol L^{-1} CaCl$_2$, pH 7.0). Bacterial growth was determined by measuring optical density at 600 nm (OD$_{600}$).
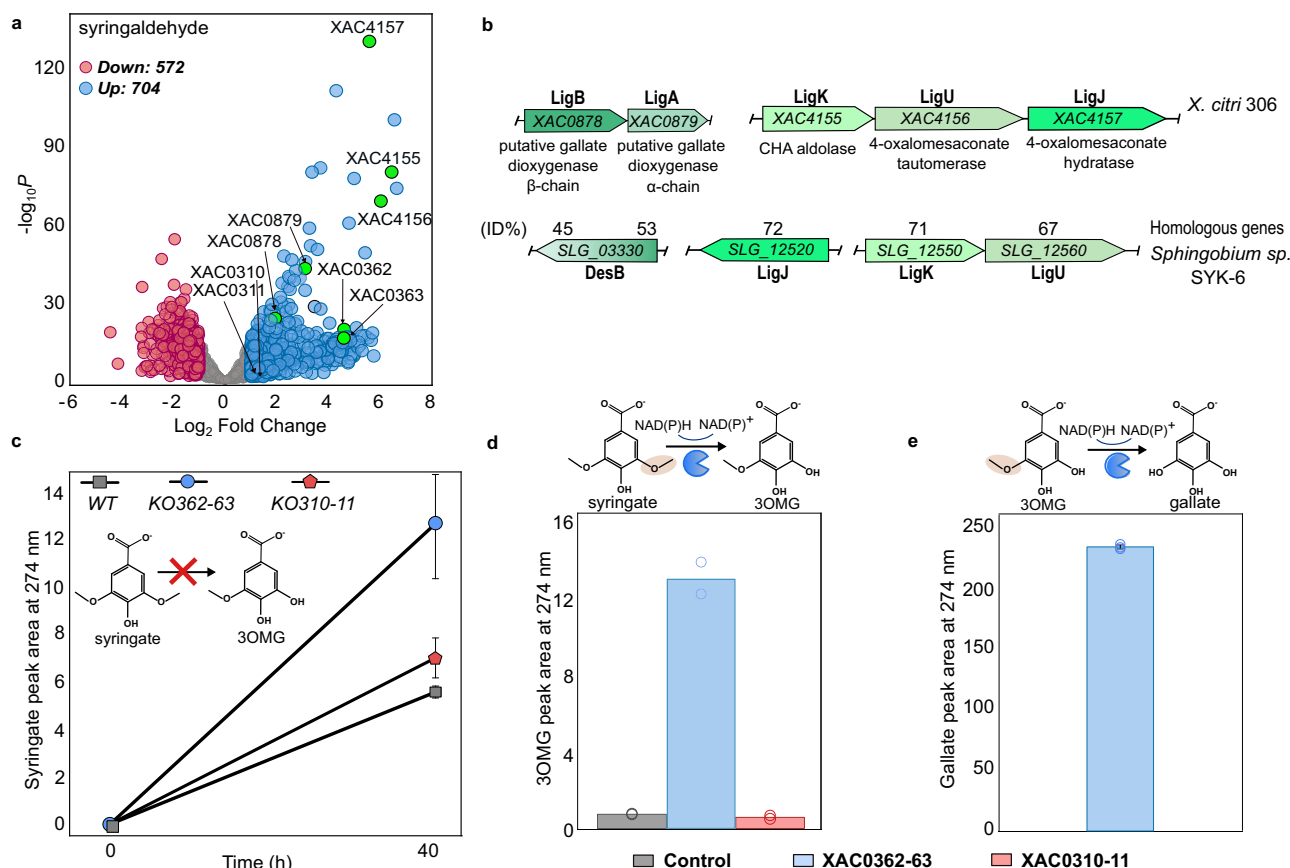
Fig. 9 | Identification of enzymes involved in the metabolism of S-type aromatic compounds. a Volcano plot of RNA-seq data highlighting in green the upregulated genes potentially related to the syringate catabolism. Differential gene expression was calculated by edgeR[76] according to likelihood ratio test within a negative binomial generalized log-linear model framework. Genes with |log2FC| >1 and adjusted *p*-value < 0.05 were assigned as DEGs. **b** Schematic representation of the genomic context of *XAC4155-4157* gene cluster and *XAC0878-0879*, showing the amino acid sequence identity with homologous enzymes of the gallate fission pathway from *Sphingobium* sp. SYK-6[12]. CHA = 4-carboxy-4-hydroxy-2-oxoadipate. **c** HPLC analysis demonstrating prominent syringate accumulation only by the Δ*XAC0362-63* strain (*KO362-63*). Data are shown as mean ± SD of *n* = 3 biological replicates. **d**, **e** Whole-cell activity assays using syringate and 3-*O*-methylgallate (3OMG) as substrates for *E. coli* BL21(DE3)-ΔslyD-pRARE2 cells expressing XAC0310-11 (red bars), XAC0362-63 (blue bars), and the negative control (empty vector – gray bars). Data are shown as mean of *n* = 2 biological replicates in d and as mean ± SD for *n* = 3 in e. Source data are provided as a source data file.

Preparation of lignin-derived compounds samples

Sugarcane bagasse was kindly provided by Isabel S/A, a Sugarcane ethanol company located in Novo Horizonte, São Paulo – Brazil. Fractions containing lignin-derived compounds were produced as previously described[62,63] (Supplementary Fig. 1). Briefly, sugarcane bagasse was subjected to an alkaline process (130 °C, 30 min, 1.5% NaOH, and 1:10 bagasse/alkaline solution ratio) in a 7.5 L reactor (Series 4580 HT, Parr) with temperature and stirring control. Next, the alkaline liquor was acidified to pH 2 with sulfuric acid (72%) to precipitate the insoluble lignin fraction[62]. After filtration, the liquid fraction containing aromatic compounds, here referred to as LDC-I, was obtained. The solid stream (precipitated lignin) was subjected to a hydrothermal depolymerization reaction at 350 °C, 165 bar, 90 min, and a lignin/water ratio of 1:50 in an inert atmosphere (N2), and an agitation of 500 rpm. This reaction was taken in a high-pressure autoclave reactor (500 mL, model 4575 A, Parr), in batch mode, up to 344 bar, 500 °C, with control over temperature, pressure, and agitation (4848 reactor controller, Parr). At the end of the reaction, the depolymerized liquid stream was subjected to a liquid-liquid extraction step with ethyl acetate (Vetec PA ACS) (1:1, 3x), and this solvent was evaporated to obtain the LDC-II fraction (bio-oil). LDC-I and LDC-II samples were characterized by UV spectroscopy, GC-MS and HPLC, as described below and in the supplementary method.

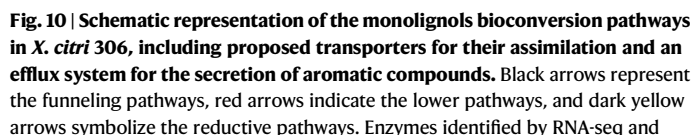## Quantification of phenolic compounds by UV spectroscopy

Concentration of total aromatic compounds in the LDC-I sample was determined by UV spectroscopy at 280 nm, on 2 M NaOH solution, pH 12, using the following equation (1)[64].

$$\text{Aromatics} \, (g \, L^{-1}) = 4.187 \times 10^{-2} * (\text{Abs}_{280nm}) - (3.279 \times 10^{-4}) * \text{dilution} \quad (1)$$

Next, it was diluted in XVM2m, filter-sterilized, and the aromatics final concentration in the medium was adjusted to 1 g L⁻¹ for *X. citri* 306 cultivations. LDC-II was solubilized on 1 mL 0.5 M NaOH solution, then diluted in minimal medium XVM2m for a final theoretical concentration of 1 g L⁻¹. After adjusting the pH to 6.7, the medium was filter-sterilized, and the concentration of aromatic compounds in the medium was estimated by 280 nm absorbance, at pH 12, according to the equation (1)[64]. Then, this medium was diluted in XVM2m to adjust the aromatics concentration to 0.3 g L⁻¹.

X. citri 306 cultivations

For growth curves analysis, *X. citri* 306 was cultured in LBON medium, 100 µg mL⁻¹ ampicillin, overnight at 30 °C and 200 rpm. Then, the harvested cells were washed once with XVM2m and inoculated for an initial OD600 of 0.05 in XVM2m supplemented with chemical standards of aromatic compounds, complex mixtures of lignin-derived compounds,

Fig. 10 | Schematic representation of the monolignols bioconversion pathways in _X. citri_ 306, including proposed transporters for their assimilation and an efflux system for the secretion of aromatic compounds. Black arrows represent the funneling pathways, red arrows indicate the lower pathways, and dark yellow arrows symbolize the reductive pathways. Enzymes identified by RNA-seq and biochemically and/or genetically validated are highlighted in green, while candidates proposed based on RNA-seq analyses, bioinformatics, and literature are shown in orange. OM outer membrane, IM inner membrane, 4HBZ 4-hydroxybenzaldehyde, 3OMG 3-_O_-methylgallate, OMA 4-oxalomesaconate.

or glucose (Supplementary Table 9). Except for LDC-I, the other conditions were also carried out in the presence of 5 mmol L^{-1} glucose as an additional carbon source. The growth was monitored in 96 well plates incubated in a SpectraMax® M3 multi-mode microplate reader (Molecular Devices) using the SoftMax Pro software or in an Infinite 200 PRO plate reader (Tecan) using the i-control software, for 48 h, at 30 °C using $n \geq 3$ biological replicates. Specific growth rates (μ) were obtained using the package growthrates as described by Petzoldt[65].

For HPLC analysis of conditions containing model compounds, *X. citri* 306 was cultivated in 125 mL flasks containing 15 mL of medium as detailed in Supplementary Table 10. The culture medium was sampled by removing 2 mL before and after specific post-inoculation time points (Supplementary Table 10), centrifuged at 4960 x *g* for 5 min to pellet the cells, and the supernatants were stored at −20 °C until the HPLC analysis described in the supplementary method.

For HPLC and GC-MS analysis of cultivations performed in minimal medium supplemented with LDC-I and LDC-II samples, *X. citri* 306 was grown in 10 mL of LBON and prepared as described above. The OD_{600} was adjusted to 0.1 in 15 mL of XVM2m supplemented with 1 g L^{-1} LDC-I or in XVM2m(G) supplemented with 0.3 g L^{-1} LDC-II. The cells were incubated at 30 °C, 200 rpm. The culture medium was sampled by removing 2 mL in 0 h and 30 h post-inoculation, centrifuged at 4960 x *g* for 5 min, and the supernatants were utilized for further analyses (LDC-I (0 h), LDC-I (30 h), LDC-II (0 h), LDC-II (30 h). The experiment was conducted with *n* = 3 biological replicates. These samples were analyzed by GC-MS and HPLC for compounds identification and quantification as described in the supplementary method. Data were analyzed using OriginPro (2021).

X. citri 306 genome mining
For the initial prediction of the metabolic pathways related to the catabolism of aromatic compounds in *X. citri* 306, we used the BLASTp tool[66] to search in the genome of this bacterium for proteins homologous to those deposited in the manually curated eLignin database[7]. As a complementary tool, we also use the metabolic maps available in the KEGG database[67] for this strain as a reference.

RNA extraction and sequencing
X. citri 306 cells grown at 30 °C, 200 rpm on 40 mL XVM2m minimal medium supplemented with different carbon sources (Supplementary Table 3) were collected at the middle exponential phase (OD_{600} = 0.05−0.1) from *n* = 4 biological replicates. Total RNA was extracted using the TRIzol/chloroform method[68]. Samples were treated with RNase-free DNaseI (Invitrogen) and RNaseOUT (Invitrogen) for 30 min, at 37 °C, and purified with the RNeasy Mini Kit (Qiagen), following manufacturer's recommendations. RNA samples concentration was determined using Nanodrop 1000 (Thermo Scientific), and their integrity was evaluated in an Agilent 2100 Bioanalyzer using the Agilent 2100 Expert Software (Agilent Technologies). The rRNA was depleted with the Ribo-Zero Plus rRNA depletion kit (Illumina Inc.). Subsequently, these samples were used to synthesize cDNA libraries with TruSeq Stranded Total RNA kit (Illumina Inc.), according to the manufacturer's recommendations. The final RNA-seq libraries were quantified via qPCR using the QIAseq Library Quant assay kit (Qiagen), and library quality was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies). Samples were pooled, and the RNA-seq was performed on an Illumina HiSeq 2500 platform equipped with HCS 2.2.68 software (SEQ facility LNBR-CNPEM, Campinas, Brazil).

RNA-seq data processing and analysis
The raw reads generated in the RNA-seq were filtered to remove adapters, primers and low-quality sequences using the fastp tool[69]. Contaminant rRNA read sequences were removed using sortmeRNA[70]. High-quality reads were mapped in the *X. citri* 306 genome (GI: 21240769) using the Bowtie2 tool[71], allowing only two mismatches and unique alignments. Next, the Samtools program[72] was used to process the alignment files, which were inspected using the Integrative Genome viewer program[73]. Reads mapped to the *X. citri* 306 genome were subjected to the featureCounts tool[74] to estimate the number of reads mapped to each transcript. The processed data were summarized and plotted using the MultiQC package[75]. Low count transcripts were removed, keeping only those that showed CPM (Counts Per Million) above 0.9, equivalent to a count of 10 to 15 reads per transcript.

Differential expression analysis was carried out using edgeR package[76] by pairwise comparisons between *X. citri* 306 grown in XVM2m supplemented with lignin-related aromatics (Supplementary Table 3) and XVM2m(G) containing 5 mmol L^{-1} glucose (reference medium). Differentially expressed genes were defined using log$_2$ Fold Change ≥ 1 (upregulated genes) or ≤ -1 (downregulated genes), and a *p*-adjusted ≤ 0.05 as thresholds. Variance analyses were conducted utilizing PCA (Principal Component Analysis) to assess data integrity and comparability. According to this analysis, one non-concordant replica from the coniferyl alcohol and one from 4-hydroxybenzaldehyde condition were excluded, and the differential expression analysis for these conditions was done with only *n* = 3 biological replicates. Functional and pathway enrichment analyses were performed separately to predict the functions of differentially expressed genes. Modular gene co-expression analysis was performed using the CEMiTool package[31] and pathway enrichment analysis using the enrichKEGG function of the clusterProfiler R package[77]. One thousand simulations were performed for each condition, and the groups were considered significantly enriched when they presented *p* ≤ 0.05. The Gene ontology (GO) enrichment analysis was performed using the clusterProfiler 3.14.3 R/ Bioconductor package and the categories were considered enriched based on hypergeometric test, implemented in the enrich function of the package[78]. Venn diagram plots were generated using the InteractiVenn web-based tool[79] (http://www.interactivenn.net/). de novo reconstruction of *molRKAB* operon from RNA-seq data was performed using the Trinity software package[80]. Comparative analysis with the reference genome (AE008923.1) was performed using pyGenomeViz package (https://moshi4.github.io/pyGenomeViz/), which was also employed for the analysis of *molRKAB* operon conservation.

Gene cloning, protein expression and purification
The open reading frames selected for biochemical assays based on the RNA-seq analysis were amplified by PCR using specific primers and cloned in the expression vector pET28a(+) or pET21b(+) using the In-Fusion® HD kit (Takara Bio) or by using restriction enzymes and DNA ligase following manufacturer's instructions (Supplementary Table 11). The constructs were confirmed by DNA sequencing and used to transform the *E. coli* strain BL21(DE3)-ΔSlyD-pRARE2 or BL21(DE3)-SHuffle®-lysY (for XAC0353 construct). The transformed cells were cultured in LB medium, 50 μg mL^{-1} kanamycin or 100 μg mL^{-1} ampicillin at 37 °C, 200 rpm until OD_{600} - 0.8. Then, the protein expression was induced by adding 0.5 mmol L^{-1} isopropyl-β-d-thiogalactopyranoside (IPTG) to the medium and incubating it at 20 °C for 16 h. Cells were harvested by centrifugation at 7,500 x *g*, at 4 °C for 30 min, resuspended on buffer as detailed in the Supplementary Table 12, and disrupted by sonication (pulses of 15 s with intervals of 30 s during 15 min, 30% amplitude). The cell extract was centrifuged at 35250 x *g*, at 4 °C for 30 min. The target proteins present in the supernatant were purified as described in the supplementary method. Protein concentration was estimated based on the absorbance of protein samples at 280 nm using the extinction coefficient calculated from their amino acid sequences using the ProtParam tool[81].

Enzyme activity screening
The activity of putative ADHs was assessed by measuring the reduction of NAD(P)$^+$ in the presence of aryl alcohols or the oxidation of NAD(P)H in the presence of aryl aldehydes (Supplementary Table 13). Putative ALDHs were evaluated for their activity over aryl aldehydes in presence of NAD(P)$^+$ (Supplementary Table 13). Substrates and cofactors were purchased as detailed in Supplementary Table 14. The enzyme activity assays were performed in 100 μL of 50 mmol L^{-1} HEPES buffer, pH 7.5, containing 0.25 mmol L^{-1} NAD(P)$^+$ or NAD(P)H and specific substrates at a final concentration of 0.25 mmol L^{-1} (Supplementary Table 14). NAD(P)H production or consumption was monitored by absorbance at 340 nm and enzyme activity was defined as described in the

supplementary method. The PobA activity was assessed in 100 μL of 50 mM Tris/sulfate buffer, pH 7.5, containing 0.5 mmol L^{-1} NADPH, 60 μmol L^{-1} FAD, and 0.5 mmol L^{-1} of either 4-hydroxybenzoate or benzoate. The reaction was initiated by adding 1 μmol L^{-1} of the PobA enzyme. A control reaction was conducted to monitor the absorption decrease at 340 nm in the absence of aryl substrates. Data were analyzed using OriginPro (2021).

Enzyme specific activity assay

The specific activity of XAC0353 and XAC0354 was accessed in 1 mL of 36 mmol L^{-1} HEPES buffer, pH 7.5, containing 0.25 mmol L^{-1} substrate, and 2 mmol L^{-1} NAD$^+$, at 30 °C for 10 min. In reactions involving coniferyl alcohol, cinnamyl alcohol, *p*-coumaryl alcohol, and sinapyl alcohol, XAC0353 was used at a concentration of 0.1 μmol L^{-1}. For the remaining aryl alcohols listed in Supplementary Table 14, a concentration of 1.2 μmol L^{-1} of XAC0353 was utilized. XAC0354 at 0.08 μmol L^{-1} was used in reactions containing aryl aldehydes (see Supplementary Table 14). Reactions were stopped by heat at 100 °C, 2 min, 650 rpm, and centrifuged (20,817 x *g*, 4 °C, 10 min). All assays were done using *n* = 3 biological replicates. The supernatant was used to measure the NADH production by HPLC as described by Sporty et al.[82], with modifications (see details in the supplementary method). Data were analyzed using OriginPro (2021).

Whole-cell activity assays using aryl aldehydes or HCAs substrates

Whole-cell assays were performed as described by García-Hidalgo et al.[49]. In short, *E. coli* BL21(DE3)-ΔSlyD-pRARE2 transformed with expression vectors containing the gene *XAC0882*, *XAC0129* or co-expressing the genes *XAC0881/XAC0883* (or with the empty vector, used as a negative control) were cultured in 5 mL of LB medium, overnight, at 30 °C and 200 rpm. Harvested cells were washed with M9 minimal medium and inoculated to an initial OD$_{600}$ of 0.8 in M9 medium containing 50 μg mL^{-1} kanamycin or 100 μg mL^{-1} ampicillin, 0.5 mmol L^{-1} IPTG, 56 mmol L^{-1} glucose, and 5 mmol L^{-1} aryl aldehydes for putative dehydrogenases (XAC0129 and XAC0882) (Supplementary Table 14) or 5 mmol L^{-1} *p*-coumarate, ferulate or sinapate for XAC0881/883. The cells were incubated at 30 °C and 200 rpm for 20 h. Samples were taken at the final time, centrifuged at 4960 x *g* for 5 min to pellet the cells, and the supernatants were analyzed as described in the HPLC section. All assays were done using *n* = 2 biological replicates.

Whole-cell activity assays for *O*-demethylases

O-demethylase whole-cell assays were performed as described by Lanfranchi et al.[83] with modifications. Briefly, the pairs XAC0310-XAC0311 or XAC0362-XAC0363 were co-expressed in *E. coli* BL21(DE3)-ΔslyD-pRARE2 in 50 mL of 2xYT medium supplemented with 1 mmol L^{-1} L-cysteine, 0.1 mg mL^{-1} FeCl$_3$, and 0.1 mg mL^{-1} FeSO$_4$ during induction with 0.5 mmol L^{-1} IPTG. After expression, the cells were centrifuged (2975 *g*, 4 °C, 10 min), washed with 20 mL of 50 mmol L^{-1} Tris-HCl buffer, pH 7.5, and resuspended in 100 mmol L^{-1} Tris-HCl, pH 7.5. The amount of buffer used for resuspension was adjusted to ensure that the OD$_{600}$ (~8.8) for the control cells (empty vector) and for the cells expressing the enzymes were the same, providing a theoretically equal number of cells across all samples. Reactions were prepared in 100 mmol L^{-1} Tris-HCl, pH 7.5, containing 0.1 mmol L^{-1} syringate, vanillate or 3OMG (final DMSO concentration 1%), 5 mmol L^{-1} DTT, 0.1 mmol L^{-1} FeSO$_4$, and 20% (v/v) of the cell stock in 2 mL Eppendorf tubes. The reactions were incubated at 30 °C, 150 rpm, for 21 h. After incubation, the reactions were stopped by heating at 90 °C for 10 min, then centrifuged (20,000 x *g*, 23 °C, 10 min). The supernatants were frozen at −20 °C before being analyzed by HPLC as described in the supplementary method. Stock solutions of substrates were prepared in 10 mmol L^{-1} DMSO and stored at 4 °C. All assays were done using *n* = 2 biological replicates.

Gene deletion

Deletion mutants were generated using established methods[84] with some modifications. In brief, DNA fragments upstream and downstream of the target gene (~1 Kbp) were amplified by PCR from *X. citri* 306 genomic DNA using the primers listed in Supplementary Tables 15 and 16. The PCR fragments were cloned into the pJET1.2/ blunt vector using The CloneJET PCR Cloning Kit (Thermo Scientific). Next, they were digested with specific restriction enzymes for sequential cloning into the suicide vector (pNPTS138) using T4 DNA ligase (Thermo Scientific). Alternatively, the PCR fragments were directly cloned into the suicide vector pNPTS138 using the commercial In-Fusion® HD kit (Takara Bio) or NEBuilder® HiFi DNA Assembly Master Mix kit (New England Biolabs). The final constructions (~2 Kbp) were confirmed by DNA sequencing. The recombinant plasmids were introduced into *X. citri* 306 by electroporation. The selection of knockout strains was performed as described in the supplementary method.

Characterization of gene knockout strains

The growth curve assay of the Δ*pobA* strain was conducted in a minimal medium XVM2m or XVM2m(G) supplemented with 4-hydroxybenzoate or benzaldehyde. The cultures were performed in sealed 96-well plates and incubated in a SPARK® multimode microplate reader (Tecan) at 30 °C with agitation. Data were collected with the SparkControl v.3 software. Each condition was replicated in at least three wells. For HPLC analysis, all mutant strains generated in this work were cultivated in 125 mL flasks containing 15 mL of medium as detailed in Supplementary Table 17. The culture medium was sampled by removing 2 mL before and after specific post-inoculation time points (Supplementary Table 17), centrifuged at 4960 x *g* for 5 min to pellet the cells, and the supernatants were stored at −20 °C until the analysis.

Phylogenetic analysis

To analyze the distribution of *molA/molB* and *calA/calB* genes in other bacterial genomes, a phylogenetic tree was constructed. The genome sequence of *X. citri* 306 along with 136 genomes were used for the analysis. The genomes were selected based on Sequence Similarity Network (SSN) analyses. Initially, individual references protein sequences of MolA (XAC0353), MolB (XAC0354), CalA (CAB69495.1) and CalB (CAA06926.1) were submitted to EFI-EST webtool[85] as seeds for SSN analyses to recover iso-functional clusters comprising the UniProt seed-related proteins sequences. The reference genome sequences of bacterial species identified on the iso-functional gene cluster harboring MolA and CalA were downloaded from the NCBI RefSeq database. This set of genome sequences was submitted to UBCG tool[86], and a phylogenetic tree was inferred based on 92 conserved single copy marker genes using maximum likelihood method implemented in RAxML with 100 bootstraps. The obtained tree was visualized in iTol (itol.embl.de). MolA/MolB and CalA/CalB related genes are highlighted in the tree according to protein sequence identity >50% and coverage >70% with references sequences as thresholds.

Statistics and reproducibility

Differentially expressed genes were evaluated according to negative binomial distribution implemented by edgeR[76] and defined using absolute log$_2$ Fold Change above 1 and adjusted *p*-values < 0.05 as thresholds. Exact adjusted *p*-values were provided in the Supplementary Data 1 file. Variance analyses were conducted utilizing PCA to assess data integrity and comparability. Outliers samples in PCA were excluded. The precise *n* number of replicate experiments are indicated in the figure legends. The investigators were not blinded to allocation during experiments and outcome assessment, since all analyses were objective in nature.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

RNA-seq data were deposited in the Gene Expression Omnibus database under accession number GSE252662. Other data generated in this study are presented in the supplementary material, supplementary data files and the source data file. Homologous sequences used for comparative purposes are found in Genbank under the accession codes CAB69495.1 [https://www.ncbi.nlm.nih.gov/protein/CAB69495.1], BAK68041.1 [https://www.ncbi.nlm.nih.gov/protein/BAK68041.1], BAH79100.1 [https://www.ncbi.nlm.nih.gov/protein/BAH79100.1], AAA75634.2 [https://www.ncbi.nlm.nih.gov/protein/AAA75634.2], BAK67507.1 [https://www.ncbi.nlm.nih.gov/protein/ BAK67507.1]. *X. citri* subsp. *citri* protein sequences, such as XAC0353, were retrieved from the KEGG database [https://www.genome.jp/entry/xac:XAC0353]. Databases used in this study: Uniprot: (https://www.uniprot.org/) eLignin (http://www.elignindatabase.com/)[7] UniProtKB/Swiss-Prot (https://www.expasy.org/resources/uniprotkb-swiss-prot) KEGG RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/) Source data are provided with this paper.

References

1. Janusz, G. et al. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol. Rev.* **41**, 941–962 (2017).
2. Khokhani, D. et al. Discovery of Plant Phenolic Compounds That Act as Type III Secretion System Inhibitors or Inducers of the Fire Blight Pathogen, Erwinia amylovora. *Appl. Environ. Microbiol.* **79**, 5424–5436 (2013).
3. Bhattacharya, A., Sood, P. & Citovsky, V. The roles of plant phenolics in defence and communication during *Agrobacterium* and *Rhizobium* infection. *Mol. Plant Pathol.* **11**, 705–719 (2010).
4. Siqueira, J. O., Nair, M. G., Hammerschmidt, R., Safir, G. R. & Putnam, A. R. Significance of phenolic compounds in plant-soil-microbial systems. *CRC Crit. Rev. Plant Sci.* **10**, 63–121 (1991).
5. Cao, B., Nagarajan, K. & Loh, K.-C. Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches. *Appl. Microbiol. Biotechnol.* **85**, 207–228 (2009).
6. Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds — from one strategy to four. *Nat. Rev. Microbiol.* **9**, 803–816 (2011).
7. Brink, D. P., Ravi, K., Lidén, G. & Gorwa-Grauslund, M. F. Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database. *Appl. Microbiol. Biotechnol.* **103**, 3979–4002 (2019).
8. Sullivan, K. P. et al. Mixed plastics waste valorization through tandem chemical oxidation and biological funneling. *Science (1979)* **378**, 207–211 (2022).
9. Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachúa, D. & Vardon, D. R. Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol.* **42**, 40–53 (2016).
10. Xu, Z., Lei, P., Zhai, R., Wen, Z. & Jin, M. Recent advances in lignin valorization with bacterial cultures: microorganisms, metabolic pathways, and bio-products. *Biotechnol. Biofuels* **12**, 32 (2019).
11. Zhao, L. et al. Biological degradation of lignin: A critical review on progress and perspectives. *Ind. Crops Prod.* **188**, 115715 (2022).
12. Kamimura, N. et al. Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism. *Environ. Microbiol. Rep.* **9**, 679–705 (2017).
13. Gu, J. et al. Bacterial transformation of lignin: key enzymes and high-value products. *Biotechnol. Biofuels Bioprod.* **17**, 2 (2024).
14. Jiménez, J. I., Miñambres, B., García, J. L. & Díaz, E. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 824–841 (2002).
15. Ravi, K., García-Hidalgo, J., Gorwa-Grauslund, M. F. & Lidén, G. Conversion of lignin model compounds by *Pseudomonas putida* KT2440 and isolates from compost. *Appl. Microbiol. Biotechnol.* **101**, 5059–5070 (2017).
16. Masai, E., Katayama, Y. & Fukuda, M. Genetic and Biochemical Investigations on Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds. *Biosci. Biotechnol. Biochem.* **71**, 1–15 (2007).
17. Masai, E., Katayama, Y., Nishikawa, S. & Fukuda, M. Characterization of *Sphingomonas paucimobilis* SYK-6 genes involved in degradation of lignin-related compounds. *J. Ind. Microbiol Biotechnol.* **23**, 364–373 (1999).
18. Jiang, W. et al. Current status, challenges and prospects for lignin valorization by using *Rhodococcus sp. Biotechnol. Adv.* **60**, 108004 (2022).
19. Morya, R., Kumar, M., Singh, S. S. & Thakur, I. S. Genomic analysis of *Burkholderia* sp. ISTR5 for biofunneling of lignin-derived compounds. *Biotechnol. Biofuels* **12**, 277 (2019).
20. Morya, R., Salvachúa, D. & Thakur, I. S. *Burkholderia*: An Untapped but Promising Bacterial Genus for the Conversion of Aromatic Compounds. *Trends Biotechnol.* **38**, 963–975 (2020).
21. Meyer, T. et al. Regulation of Hydroxycinnamic Acid Degradation Drives *Agrobacterium fabrum* Lifestyles. *Mol. Plant-Microbe Interact.* **31**, 814–822 (2018).
22. Broek, A. V. & Vanderleyden, J. The role of bacterial motility, chemotaxis, and attachment in bacteria plant interactions. *Mol. Plant-Microbe Interact.* **8**, 800–810 (1995).
23. Vieira, P. S. et al. Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* **12**, 4049 (2021).
24. 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. *N. Phytologist* **198**, 899–915 (2013).
25. Giuseppe, P. O., Bonfim, I. M. & Murakami, M. T. Enzymatic systems for carbohydrate utilization and biosynthesis in *Xanthomonas* and their role in pathogenesis and tissue specificity. *Essays Biochem.* **67**, 455–470 (2023).
26. William, F. & Mahadevan, A. Degradation of aromatic compounds by *Xanthomonas* species / Abbau aromatischer Verbindungen durch *Xanthomonas*-Arten. *J. Plant Dis. Prot.* **87**, 738–744 (1980).
27. Chen, B. et al. The phytopathogen *Xanthomonas campestris* utilizes the divergently transcribed *pobA* / *pobR* locus for 4-hydroxybenzoic acid recognition and degradation to promote virulence. *Mol. Microbiol.* **114**, 870–886 (2020).
28. Wang, J.-Y. et al. A functional 4-hydroxybenzoate degradation pathway in the phytopathogen *Xanthomonas campestris* is required for full pathogenicity. *Sci. Rep.* **5**, 18456 (2015).
29. Chen, B. et al. The phytopathogen *Xanthomonas campestris* scavenges hydroxycinnamic acids in planta via the *hca* cluster to increase virulence on its host plant. *Phytopathol. Res.* **4**, 12 (2022).
30. Daurelio, L. D. et al. Novel insights into the Citrus sinensis nonhost response suggest photosynthesis decline, abiotic stress networks and secondary metabolism modifications. *Funct. Plant Biol.* **42**, 758 (2015).
31. Russo, P. S. T. et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinforma.* **19**, 56 (2018).
32. Achterholt, S., Priefert, H. & Steinbüchel, A. Purification and Characterization of the Coniferyl Aldehyde Dehydrogenase from *Pseudomonas* sp. Strain HR199 and Molecular Characterization of the Gene. *J. Bacteriol.* **180**, 4387–4391 (1998).

33. Overhage, J., Steinbüchel, A. & Priefert, H. Biotransformation of Eugenol to Ferulic Acid by a Recombinant Strain of *Ralstonia eutropha* H16. *Appl Environ. Microbiol.* **68**, 4315–4321 (2002).

34. Song, Y.-N., Shibuya, M., Ebizuka, Y. & Sankawa, U. Identification of Plant Factors Inducing Virulence Gene Expression in Agrobacterium tumefaciens. *Chem. Pharm. Bull. (Tokyo)* **39**, 2347–2350 (1991).

35. Westphal, A. H., Tischler, D. & van Berkel, W. J. H. Natural diversity of FAD-dependent 4-hydroxybenzoate hydroxylases. *Arch. Biochem Biophys.* **702**, 108820 (2021).

36. Entsch, B. & Van Berkel, W. J. H. Structure and mechanism of para-hydroxybenzoate hydroxylase. *FASEB J.* **9**, 476–483 (1995).

37. Venkatesagowda, B. & Dekker, R. F. H. Microbial demethylation of lignin: Evidence of enzymes participating in the removal of methyl/methoxyl groups. *Enzym. Micro. Technol.* **147**, 109780 (2021).

38. Wolf, M. E., Hinchen, D. J., DuBois, J. L., McGeehan, J. E. & Eltis, L. D. Cytochromes P450 in the biocatalytic valorization of lignin. *Curr. Opin. Biotechnol.* **73**, 43–50 (2022).

39. Mallinson, S. J. B. et al. A promiscuous cytochrome P450 aromatic O-demethylase for lignin bioconversion. *Nat. Commun.* **9**, 2487 (2018).

40. Kamimura, N. et al. A bacterial aromatic aldehyde dehydrogenase critical for the efficient catabolism of syringaldehyde. *Sci. Rep.* **7**, 44422 (2017).

41. Abe, T., Masai, E., Miyauchi, K., Katayama, Y. & Fukuda, M. A Tetrahydrofolate-Dependent *O*-Demethylase, LigM, Is Crucial for Catabolism of Vanillate and Syringate in *Sphingomonas paucimobilis* SYK-6. *J. Bacteriol.* **187**, 2030–2037 (2005).

42. Nogales, J., Canales, Á., Jiménez-Barbero, J., García, J. L. & Díaz, E. Molecular Characterization of the Gallate Dioxygenase from *Pseudomonas putida* KT2440. *J. Biol. Chem.* **280**, 35382–35390 (2005).

43. Uchendu, S. N. et al. Identifying metabolic pathway intermediates that modulate the gallate dioxygenase (DesB) from *Sphingobium* sp. strain SYK-6. *Process Biochem.* **102**, 408–416 (2021).

44. Overhage, J., Priefert, H. & Steinbüchel, A. Biochemical and Genetic Analyses of Ferulic Acid Catabolism in *Pseudomonas* sp. Strain HR199. *Appl Environ. Microbiol* **65**, 4837–4847 (1999).

45. Belchik, S. M., Schaeffer, S. M., Hasenoehrl, S. & Xun, L. A. β-barrel outer membrane protein facilitates cellular uptake of polychlorophenols in *Cupriavidus necator*. *Biodegradation* **21**, 431–439 (2010).

46. Steinbichel, Alexander, Priefert, Horst & Rabenhorst, Jirgen *Synthetic enzymes for the production of coniferyl alcohol, coniferylaldehyde, ferulic acid, vanillin and vanillic acid and their use.*, (2003).

47. Peleg, H., Naim, M., Rouseff, R. L. & Zehavi, U. Distribution of bound and free phenolic acids in oranges (*Citrus sinensis*) and Grapefruits (*Citrus paradisi*). *J. Sci. Food Agric* **57**, 417–426 (1991).

48. Liu, S. et al. Review of phytochemical and nutritional characteristics and food applications of *Citrus* L. fruits. *Front Nutr.* **9**, 968604 (2022).

49. García-Hidalgo, J. et al. Vanillin Production in *Pseudomonas*: Whole-Genome Sequencing of *Pseudomonas* sp. Strain 9.1 and Reannotation of *Pseudomonas putida* CalA as a Vanillin Reductase. *Appl Environ. Microbiol* **86**, e02442–19 (2020).

50. Zhou, P., Khushk, I., Gao, Q. & Bao, J. Tolerance and transcriptional analysis of *Corynebacterium glutamicum* on biotransformation of toxic furaldehyde and benzaldehyde inhibitory compounds. *J. Ind. Microbiol Biotechnol.* **46**, 951–963 (2019).

51. Zhou, H., Xu, Z., Cai, C., Li, J. & Jin, M. Deciphering the metabolic distribution of vanillin in *Rhodococcus opacus* during lignin valorization. *Bioresour. Technol.* **347**, 126348 (2022).

52. Weiland, F., Barton, N., Kohlstedt, M., Becker, J. & Wittmann, C. Systems metabolic engineering upgrades *Corynebacterium glutamicum* to high-efficiency cis, cis-muconic acid production from lignin-based aromatics. *Metab. Eng.* **75**, 153–169 (2023).

53. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. in *Plant Bioinformatics* 89–112 (Humana Press, Totowa, 2007).

54. Wang, H. et al. Functions of aldehyde reductases from *Saccharomyces cerevisiae* in detoxification of aldehyde inhibitors and their biotechnological applications. *Appl Microbiol Biotechnol.* **102**, 10439–10456 (2018).

55. Chen, Z., Sun, X., Li, Y., Yan, Y. & Yuan, Q. Metabolic engineering of Escherichia coli for microbial synthesis of monolignols. *Metab. Eng.* **39**, 102–109 (2017).

56. Lv, Y. et al. Improving bioconversion of eugenol to coniferyl alcohol by in situ eliminating harmful H_2O_2. *Bioresour. Technol.* **267**, 578–583 (2018).

57. Chen, Z. et al. Establishing an Artificial Pathway for De Novo Biosynthesis of Vanillyl Alcohol in *Escherichia coli*. *ACS Synth. Biol.* **6**, 1784–1792 (2017).

58. Abid, M., Khan, M. A. & Wahid, A. Screening and determination of phenolics in relation to resistance against citrus canker. *Pak. J. Phytopathol.* **20**, 109–116 (2008).

59. González, J. J. et al. Transcriptional changes involved in kumquat (*Fortunella spp*) defense response to *Xanthomonas citri* subsp. *citri* in early stages of infection. *Physiol. Mol. Plant Pathol.* **116**, 101729 (2021). Giraldo.

60. Lin, C.-Y. et al. Engineering sorghum for higher 4-hydroxybenzoic acid content. *Metab. Eng. Commun.* **15**, e00207 (2022).

61. Hao, Z. et al. Expression of a bacterial 3-dehydroshikimate dehydratase (QsuB) reduces lignin and improves biomass saccharification efficiency in switchgrass (Panicum virgatum L.). *BMC Plant Biol.* **21**, 56 (2021).

62. de Menezes, F. F. et al. Alkaline Pretreatment Severity Leads to Different Lignin Applications in Sugar Cane Biorefineries. *ACS Sustain Chem. Eng.* **5**, 5702–5712 (2017).

63. de Menezes, F. F. et al. Exploring the compatibility between hydrothermal depolymerization of alkaline lignin from sugarcane bagasse and metabolization of the aromatics by bacteria. *Int J. Biol. Macromol.* **223**, 223–230 (2022).

64. Rocha, G. J. et al. Influence of mixed sugarcane bagasse samples evaluated by elemental and physical–chemical composition. *Ind. Crops Prod.* **64**, 52–58 (2015).

65. Petzoldt, Thomas *Estimation of Growth Rates with Package growthrates*. Preprint at https://tpetzoldt.github.io/growthrates/doc/Introduction.html (2022).

66. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

67. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

68. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem* **162**, 156–159 (1987).

69. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

70. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

71. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

72. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

73. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform* **14**, 178–192 (2013).

74. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

75. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

76. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

77. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

78. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).

79. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinforma.* **16**, 169 (2015).

80. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

81. Gasteiger, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (Humana Press, Totowa, 2005).

82. Sporty, J. L. et al. Single sample extraction protocol for the quantification of NAD and NADH redox states in *Saccharomyces cerevisiae*. *J. Sep Sci.* **31**, 3202–3211 (2008).

83. Lanfranchi, E., Trajković, M., Barta, K., de Vries, J. G. & Janssen, D. B. Exploring the Selective Demethylation of Aryl Methyl Ethers with a *Pseudomonas* Rieske Monooxygenase. *ChemBioChem* **20**, 118–125 (2019).

84. Tófoli de Araújo, F. et al. Structural and Physiological Analyses of the Alkanesulphonate-Binding Protein (SsuA) of the *Citrus* Pathogen *Xanthomonas citri*. *PLoS One* **8**, e80083 (2013).

85. Gerlt, J. A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et. Biophysica Acta (BBA) - Proteins Proteom.* **1854**, 1019–1037 (2015).

86. Na, S.-I. et al. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol.* **56**, 280–285 (2018).

Acknowledgements

Author contributions

D.B.M. performed *X. citri* 306 genome mining, *X. citri* 306 cultivations, RNA extraction, *molA* and *molB* gene knockout, Δ*molA*, Δ*molB*, Δ*pobA* and Δ*pcaHG* strains characterization, XAC0881, XAC0883 and dehydrogenases cloning, activity screening, and whole cell assays. A.J.V.C.B. performed XAC3477 purification, MolA and MolB cloning, expression, purification, and enzyme assays. A.R.L. performed XAC0129 cloning and ROs-type *O*-demethylases cloning, whole cell assays, gene knockout and strains characterization. L.D.W., M.S.C. and M.C.G. designed and performed HPLC assays. D.B.M., D.A.A.P., J.M.J., G.F.P. and P.O.G. designed and performed RNA-seq assays and analyses. F.F.M., J.A.A. and R.P. designed and performed GC-MS assays. J.M.J. and G.F.P. performed XAC0351-54 operon conservation analysis and phylogenetic analysis. F.M.K. performed *pobA* and *pcaHG* gene knockout. F.F.M., and G.J.M.R. designed and performed lignin depolymerization reactions. P.O.G. coordinated the work. D.B.M. and P.O.G. wrote the manuscript. All authors contributed to data analysis and manuscript revision.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.