Widespread distribution of *hmf* genes in Proteobacteria reveals key enzymes for 5-hydroxymethylfurfural conversion

Raúl A. Donoso a,b, Fabián González-Toro a, Danilo Pérez-Pantoja a,⇑

a Programa Institucional de Fomento a la Investigación, Desarrollo e Innovación (PIDi), Universidad Tecnológica Metropolitana, Santiago, Chile

b Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile

A R T I C L E   I N F O

Article history:
Received 7 November 2020
Received in revised form 5 April 2021
Accepted 5 April 2021
Available online 16 April 2021

Keywords:
HMF
Furans
Biodegradation
Proteobacteria
Metabolic engineering

A B S T R A C T

Furans represent a class of promising chemicals, since they constitute valuable intermediates in conversion of biomass into sustainable products intended to replace petroleum-derivatives. Conversely, generation of furfural and 5-hydroxymethylfurfural (HMF) as by-products in lignocellulosic hydrolysates is undesirable due its inhibitory effect over fermentative microorganisms. Therefore, the search for furans-metabolizing bacteria has gained increasing attention since they are valuable tools to solve these challenging issues. A few bacterial species have been described at genetic level, leading to a proposed HMF pathway encoded by a set of genes termed *hnfpfsf*, although some enzymatic functions are still elusive. In this work we performed a genomic analysis of major subunits of furoyl-CoA dehydrogenase orthologues, revealing that the furoic acid catabolic route, key intermediate in HMF biodegradation, is widespread in proteobacterial species. Additionally, presence/absence profiles of *hnfpfsf* genes in selected proteobacterial strains suggest parallel and/or complementary roles of enzymes with previously unclear function that could be key in HMF conversion. The furans utilization pattern of selected strains harboring different *hnfpfsf* gene sets provided additional support for bioinformatic predictions of the relevance of some enzymes. On the other hand, at least three different types of transporter systems are clustered with *hnfpfsf* genes, whose presence is mutually exclusive, suggesting a core and parallel role in furans transport in Proteobacteria. This study expands the number of bacteria that could be recruited in biotechnological processes for furans biodetoxification and predicts a core set of genes required to establish a functional HMF pathway in heterologous hosts for metabolic engineering endeavors.

1. Introduction

The sustainable production of bio-based chemicals is greatly reliant on the development of viable biorefineries to use different biomass resources for their conversion into diversified products [1]. These bioproducts have the potential to replace those produced by petroleum refineries [1]. Furans represent one of the most promising classes of intermediates in the conversion of biomass into bio-based products since they constitute sustainable alternatives to various petroleum-derived chemicals, having properties analogous to and even exceeding those derived from fossil hydrocarbons [2,3]. Particularly, 5-hydroxymethylfurfural (HMF) has been termed as the “Sleeping Giant” of sustainable chemistry because of its vast synthetic potential, and it has been broadly accepted by the scientific community as a key element to bridge the gap from a fossil-based economy to a sustainable manufacturing system [2,4]; being the focus of nearly 1000 papers published per year [2] and reaching a global market of 56 million US$ in 2019 [5]. HMF can be produced from the dehydration of hexoses (mainly fructose) or via acid-catalyzed hydrolysis/dehydration of cellulose [6]. The latter synthetic route is the second aspect of HMF chemistry that is relevant for biotechnology industry, this is due its production as by-product during the acid-based pretreatment of lignocellulose to allow the subsequent digestion of cellulose for biofuels fermentation [6]. The generation of furans in lignocellulosic hydrolysates is unwanted due its highly inhibitory effect over fermentative microorganisms, and consequently HMF is included along with furfural in the so-called “lignocellulose-derived inhibitors” (LDIs) group of noxious compounds whose adverse effects include decreased viability, ethanol yield, and productivity [7,8]. In particular, at levels found in hydrolysates, HMF can produce a reduction of approximately 33% in growth rate of an ethanologenic bacterium such as *Zymomonas mobilis* [9], and

https://doi.org/10.1016/j.csbj.2021.04.017

2001-0370© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
in mixtures with other LDIs is able to severely affect ethanol yield and productivity by 90 and 85% respectively, when tested in the fermentative yeast *Scheffersomyces stipitis* [10].

Several approaches have been explored to overcome the negative effect of furan inhibitors, including decreasing their generation by adjusting pretreatment conditions, developing more tolerant ethanologenic strains, and detoxifying the hydrolysates [7,8,11]. Regarding this latter approach, biological detoxification, using a microorganism to remove inhibitors from the hydrolysate, is an attractive alternative due minimal generation of waste streams, no need for chemical inputs, and the feasibility to be performed directly in the fermentation vessel avoiding an extra process step [7,8,11]. The cornerstone of this approach is to use a very proficient microorganism having abilities to catabolize multiple furans without metabolization of sugars intended for fermentation [12–15]. Altogether, the requirement of furans-metabolizing strains both for biological detoxification and for biocatalysis aimed to obtain fine products have promoted the study of furans biodegradation over the last years [16–18].

A limited number of bacterial species owning the ability to use furfural or HMF as sole carbon sources have been characterized at genetic level, with *Pseudomonas putida* Fu1 [19], *Cupriavidus basilensis* HMF14 [20], and *Pseudomonas putida* ALS1267 [21] being the best characterized examples. The pioneering work of Nichols and Mertens provided the first pieces of evidence about furans biodegradation in *P. putida* strain Fu1 by identification of catabolic and regulatory *psf* genes [19]. Later, Koopman et al. [20] performed the genetic characterization of HMF and furfural degradation in *C. basilensis* HMF14, both converging to 2-furoic acid (FA) to be converted to 2-oxoglutarate using coenzyme A intermediates and enzymes encoded by *hmfABC* genes (Fig. 1). In the case of HMF catabolism, the presence of *hmfFGH* genes is also required since they are involved in the upper part of the degradative route, supporting the conversion of HMF into FA (Fig. 1) [18,20]. Specifically, *hmfFG* genes encode a decarboxylase, able to catalyze conversion of key intermediate 2,5-furandicarboxylic acid (FDCA) to FA, whilst *hmfH* encodes a FAD-dependent oxidoreductase putatively involved in two consecutive oxidations of 5-hydroxymethyl-2-fural ancocarboxylic acid (HMFA) to generate FDCA [18,20]. No specific aldehyde dehydrogenases have been identified for initial HMF oxidation into HMFCA, suggesting that such activity is related to broad-specificity dehydrogenases protecting cells against toxic aldehydes [18,20,22]. Recently, the genes involved in HMF catabolism have been characterized in *P. putida* ALS1267, which carries all *hmf* genes previously described in strain HMF14 except for *hmfH*, suggesting that another gene would replace this FAD-dependent oxidoreductase [21]. Interestingly, two additional genes are present in *hmf* operon from *P. putida* ALS1267, which were also previously found in *P. putida* Fu1 and termed *psfG* and *psfA*, encoding a putative short-chain alcohol dehydrogenase and a putative NAD-dependent aldehyde dehydrogenase, respectively [19]. These enzymatic functions have been suggested to be related to the oxidation of furfuryl alcohol/furfural and HMF alcohol/HMF [19,21], but additional support has not been provided.

The interest to establish a furans detoxification pathway in fermentative microorganisms has driven metabolic engineering endeavors for functional expression of HMF route in genetically tractable strains [23–26]. Two studies focused on transferring the entire HMF pathway from *P. putida* ALS1267 [21] and from a previously unrecognized furans-degrading *Paraburkholderia phytotrichorum* PsJN strain [24], to industrially relevant *P. putida* strains. These efforts have not been completely successful since the furans-degrading phenotype was only partially achieved as revealed by lack of growth on HMF or a very slow proliferation rate by the engineered strains [21,24]; most probably because the selection of catabolic genes was incomplete, or their expression was inadequate.

Considering that metabolic engineering has as a central goal to identify gene targets for the optimization of a metabolic phenotype [27], we focus on genome analysis of furans biodegradation to pro-
vide additional insights about genetic requirements for HMF conversion. In this work we performed an extensive survey on the presence of hmfA/psf genes in proteobacterial genomes revealing a previously unrecognized distribution of the HMF pathway, broadly expanding the diversity of species suitable for furans conversion. In addition, the presence-absence pattern of some genes suggests redundant or complementary roles that should be considered in metabolic engineering endeavors towards biological detoxification of furans and/or biocatalytic production of high-value intermediates.

2. Material and methods

2.1. Database survey

Protein similarity searches were performed using the blastP algorithm from NCBI and IMG/M databases using default parameters [28,29]. The HmfA sequences were retrieved from non-redundant protein database of GenBank [30]. Only proteins displaying at least 45% amino acid identity with the hmfA-encoded protein of C. basilensis HMFI4 [20] were considered for analysis. Proteobacterial genomes that harbor hmfA homologues were downloaded from NCBI database for identification of additional catabolic and complementary functions. Gene products of lower (hmfBCDE) and upper (hmfFGH and psfAG) pathways, transcriptional regulators (AraC- and LysR-type), and transporters (hmfTT) with verified or putative functions located in the hmf gene clusters of C. basilensis HMFI4, P. putida ALS1267, and P. putida FuI1 [19–21] were used as bait for homologue searches in genomes previously selected using BLAST software [28] on a local computer. Identification of conserved domains within a protein was performed using the Conserved Domains Database from the NCBI website [31].

2.2. Phylogenetic reconstructions

Evolutionary relationships from the concatenated alignment of the single-copy hmfA/BCDE genes were inferred by IQ-TREE web server tools (http://iqtrees.cibiv.univie.ac.at) designed for estimate maximum-likelihood phylogenies [32] employing ModelFinder as model-selection method [33] and Ultrafast Bootstrap Approximation as bootstrap approach [34] with the -m TEST, -bb 1000 strategy [35]. Visualization and edition of phylogenetic trees were performed by the Interactive Tree Of Life (ITOL) online tool (https://itol.embl.de/) [36].

2.3. Bacterial strains and growth conditions

The bacterial strains assayed in this study were Azospirillum oryzae COCC [37], Cupriavidus basilensis DSM 11853 [38], Cupriavidus pinatubonensis JSMT134 [39], Paraburkholderia phytofirmans PsjN [40], Paraburkholderia xenovorans LB400 [41], Pseudomonas mendocina VN230 [42], Pseudomonas sp. ALS1131 [43], and Pseudomonas sp. ALS1279 [43]. All strains were grown in mineral salts medium (14 g/L Na2HPO4·12H2O, 2 g/L KH2PO4, 50 mg/L Ca(NO3)2·4H2O, 1 g/L (NH4)2SO4, 200 mg/L MgSO4·7H2O, 278 µg/L FeSO4·7H2O, 70 µg/L ZnCl2, 100 µg/L MnCl2·4H2O, 62 µg/L H3BO3, 190 µg/L CoCl2·6H2O, 17 µg/L CuCl2·2H2O, 24 µg/L NiCl2·6H2O, 36 µg/L Na2MoO4·2H2O), supplemented with 2.5 mM HMF, HMFCA, 5-formylfuran carboxylic acid (FFCA), FDCA or FA as sole carbon and energy source. The cultures were incubated in a 96-well microplate (Thermo Fisher Scientific, Rochester, NY, USA) at 30 °C, and the optical density at 600 nm (OD600) was measured with a Synergy HTX multimode plate reader (BioTek, Winooski, VT, USA). Cell cultures were inoculated with 100-fold dilutions of overnight cultures grown on R2A broth. At least two biological replicates were performed for each growth measurement.

2.4. Chemicals

HMF, FFCA and FA were purchased from Sigma-Aldrich (Steinheim, Germany). HMFCA was purchased from Cayman Chemical Company (Michigan, USA). FDCA was purchased from Alfa Aesar (Massachusetts, USA).

3. Results and discussion

3.1. Furans biodegradation genes are widespread in α-, β- and γ- lineages of the phylum Proteobacteria

In order to evaluate the presence of furans biodegradative genes in proteobacterial genomes we chose as gene marker the HmfA product corresponding to the major subunit of furoyl-CoA dehydrogenase, a key enzyme in the lower route responsible for 2-furoyl-CoA metabolism (Fig. 1). Using HmfA from C. basilensis HMFI4 [20] as query (1015 aa), we conducted a restricted search for Proteobacteria in the non-redundant protein sequences database of GenBank as of September 2020, identifying nearly 900 HmfA-like sequences displaying at least 45% amino acid identity in strains that belong to α-, β- and γ-Proteobacteria lineages. We selected 98 bacterial species with genome sequences available to include at least one member of each genus harboring HmfA homologues in order to examine the presence of additional hmf genes (Fig. 1). We also include the genomes of three additional Pseudomonas species (ALS1131, ALS1267 and ALS1279) with verified furans-degrading abilities [43,44] collecting a total of 101 strains. Nevertheless, two bacterial species harboring HmfA, representing Inquilinus (GCA_002195995.1) and Acidisphaera (GCA_000964365.1) genus of the α-Proteobacteria lineage (order Rhodospirillales) were subsequently discarded due to low quality genome assembly. From the 99 Proteobacterial genomes analyzed (Table S1), we found that 50 belong to α-Proteobacteria class subdivided in order Rhizobiales (30), Rhodobacterales (14), Rhodospirillales (5) and Sphingomonadales (1), representing 14 different families (Fig. S1); 35 belong to β-Proteobacteria class subdivided in order Burkholderiales (33) and Rhodocyclales (2), representing 6 different families (Fig. S1); and 14 belong to γ-Proteobacteria class subdivided in order Alteromonadales (1), Chromatiales (1), Oceanospirillales (5), Pseudomonadales (4) (including three bona-fide degrading strains added), Salinisphaerales (1) and Xanthomonadales (2), representing 8 different families (Fig. S1). These strains were isolated from diverse locations such as oceans, plants, freshwater, groundwater, sediments, soils, activated sludge, wastewater, saline springs and human host, among other environments (Table S1). Additionally, we detected the hmfA gene in metagenome-assembled genomes belonging to Candidimonas (TAMB5542.1), Comamonas (MPT12907.1), Confluentimicrobium (MAQ45687.1), Filomicrobium (MAH44436.1), Kastaia (ODT20617.1), Lautropia (ODS93822.1), Methylibium (KZN21457.1) and Verminephrobacter (MBA4060343.1) genera, obtained from a wide diversity of environmental sources. These genomes not represented by isolated bacterial species expand the range of proteobacterial genera putatively harboring the furans-degrading pathway. Then, we analyzed the presence of remaining lower pathway genes (hmfBCDE) in selected strains, observing that all bacterial species have the full set of genes required for metabolization of FA, the central intermediate in fur-
ans catabolism (Fig. 1; Table S1). These results would indicate a ubiquitous presence of furans in the environment and their relevance as natural substrates.

In order to establish phylogenetic relationship among hmf genes of selected strains we constructed a dendogram of concatenated amino acid sequences of lower pathway genes. The resulting HmfABCDE tree showed a strict partition between Proteobacteria lineages with few exceptions (Fig. 2). Interestingly, γ-Proteobacteria Pseudoxanthomonas compositi GSS15 isolated from compost [45] and Luteimonas huabeiensis HB-2 isolated from stratum water [46], both belonging to the family Xanthomonadaceae, are inserted in the clade of sequences belonging to α-Proteobacteria (Fig. 2), suggesting lateral gene transfer of lower pathway genes. Furthermore, vast majority of strains (85) harbor the canonical hmfABCDE gene order (Fig. 3; Table S1), indicating that this organization is highly conserved to that reported by Wierckx et al. [18]. The second most common organization was hmfABCFGE, which is restricted to β-Proteobacteria (11) and includes the insertion of hmfFG genes of upper pathway (Fig. 3; Table S1). Only Rhodoplanes sp. Z2-YC6860 and Noviherbaspirillum denitrificans TSA40 have lower pathway genes in different gene contexts (Fig. 3), raising questions about its functionality in these strains. Remarkably, β-Proteobacteria strains Acidovorax sp. KKS102, Bordetella flabilis AU10664 and Paracandidimonas soli DSM 100048 harbor two copies of hmfABCDE gene cluster (Fig. 3), suggesting functional redundancy of lower pathway in these bacteria. In the case of P. soli DSM 100048 both copies are very close related indicating a recent event of gene duplication. From the total number of analyzed strains, 44 species harbor exclusively lower pathway genes and not those related to upper route for HMF, suggesting that FA is per se a relevant growth substrate or it is metabolized as central intermediate for channeling of furfural by non-specific dehydrogenases. On the other hand, 55 selected strains carry extra genes related to upper pathway

![Fig. 2. Evolutionary relationships among concatenated HmfABCDE protein homologues identified in proteobacterial genomes analyzed in this work. Maximum likelihood topology provided by IQ-TREE [32] and based on sequence alignments calculated using MAFFT [35] is shown with SH-like approximate likelihood ratio support values (n = 1000) given at each node (values > 50% are shown). The substitution model selected was LG + F + I + G4 as determined by ModelFinder [33]. Blue, α-proteobacteria; orange, β-proteobacteria; green, γ-proteobacteria. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
3.2. Additional catabolic enzymes could be part of the HMF pathway in Proteobacteria

A key step in HMF degradation is played by hmfFG genes, which encode an enzymatic activity able to catalyze FDCA decarboxylation to FA (Fig. 1). Since HmfG is homologous to UbiX, which synthesizes the prenylated flavin mononucleotide cofactor required by UbID family of enzymes [47,48], apparently the key catalytic function in FDCA decarboxylation is performed by HmfF, which belongs to UbiD family [49]. We detected 55 bacterial genera out of 99 selected representative strains that harbor hmfFG genes, in which 18 belong to α-Proteobacteria (36% of selected strains belonging to this lineage), 28 are classified as β-Proteobacteria (80%), and 9 are included in the γ-Proteobacteria lineage (64%) (Fig. 4), suggesting broader spectrum of furan compounds degraded by these species.

An enzyme playing a crucial role in the upper pathway of furans biodegradation is the HMF oxidoreductase encoded by hmfH, previously described in C. basilensis HMF14, which has been extensively studied for its ability to convert HMF into FDCA in heterologous strains P. putida S12 and Raoultella ornithinolytica BF60, turning a valuable tool for production of this high-value precursor [22,26,50–52]. HmfH is a member of the glucose-methanol-choline (GMC) oxidoreductase family and mainly converts HMFCA into FFCA, and secondarily it would be also involved in the subsequent oxidation into FDCA [18,26]. However, a role for unspecified aldehyde dehydrogenases catalyzing oxidation of FFCA into FDCA during heterologous expression of hmf genes in P. putida S12 has been also proposed [22]. In the selected strains harboring the lower route for furans we found overall 23 species that harbor hmfH gene.
mainly α- (7) and β-Proteobacteria (14) (Fig. 4). We only detected this oxidoreductase-encoding gene in two γ-Proteobacteria, P. composti CSS15 and L. huabeiensis HB-2 (Table S1). Notably, the hmfABCDEFG genes of both strains were previously clustered into the α-Proteobacteria clade in hmfABCDEFG tree (Fig. 2), strongly suggesting that these clusters have been acquired by lateral gene transfers, and consequently pinpointing that hmfH gene would not be genuinely associated to hmf clusters originated in the γ-Proteobacteria. This indicates that in most HMF-degrading γ-Proteobacteria the consecutive oxidations of HMFCA to generate FDCA should be performed by enzymes other than HmfH homologues. Notably, in the search for alternative enzymes to be recruited for biocatalytic FDCA production a remote HmfH homologue has been identified in Methyllovorus sp. MP688 (46% aa identity) and termed HmfO [53]. Due to its biocatalytic potential, the properties of HmfO have been comprehensively studied [52,54–57]. It should be noted that most probably HMFCA or FFCA are not physiological substrates for this enzyme since no hmfABCDEFG genes could be identified by us in the genome of Methyllovorus sp. MP688 or related Methyllovorus species, excluding a role for HmfO in a catabolic route for HMF.

Interestingly, the psfA gene encoding a putative aldehyde dehydrogenase previously identified in P. putida strains Fu1 and ALS1267 [19,21] was additionally detected in 42 genomes out of 55 strains harboring hmfABCDEFG genes, including α- (9), β- (24) and γ-Proteobacteria (9) (Fig. 4; Fig. S3), suggesting a relevant role of this enzyme in the upper pathway for furans. A homologue of this putative aldehyde dehydrogenase (63% aa identity with psfA) was previously detected in close proximity to hmf genes of C. basiliscum HMF14 and termed Adh [18]. Notably, during the establishment of a whole cell biocatalyst system by heterologous expression of hmfH gene in P. putida S122 [26], it was demonstrated that FDCA production was improved by achieving a high and constant conversion rate by co-expression of adh and a transporter gene termed hmfT1 (see 3.4 section), avoiding transient accumulation of FFCA detected in presence of hmfH and hmfT1 [22]. These results suggest that a PsfA-like aldehyde dehydrogenase would be involved in conversion of FFCA into FDCA, complementing the activity of HmfH over HMFCA. Accordingly, we detected that 16 out of 23 strains containing the HmfH-encoding gene also own a psfA gene in its genome (Fig. 4), supporting importance of aldehyde dehydrogenase activity of PsfA (most probably as FFCA dehydrogenase) in HMF pathway (Fig. 1). It should be noted that in 7 strains harboring a hmfH gene (5 α- and 2 β-Proteobacteria) a psfA gene could not be found (Fig. 4; Table S1), suggesting that if these strains are able to grow in HMF as carbon source they should recruit a different dehydrogenase activity to replace for PsfA. Most probably, the PsfA aldehyde dehydrogenase has an underestimated role in the HMF degradation pathway (Fig. 1), masked by the pool of endogenous aldehyde dehydrogenases that protects cells against aldehydes toxicity.

Another gene originally identified in Pseudomonas strains Fu1 and ALS1267 is psfG, encoding a putative short-chain alcohol dehydrogenase putatively involved in furfuryl(HM)F alcohol oxidation [19,21]. The psfG gene was found in 19 species comprising α- (3), β- (9) and γ-Proteobacteria (7), all of them simultaneously harboring hmfFG and psfA genes (Fig. 4). A comparison of the amino acid sequence of psfG (249 aa) from ALS1267 strain [21] in the UniProtKB/Swiss-Prot database showed a 39% sequence identity with (S)-1-Phenylethanol dehydrogenase from Aromatoleum aromaticum EbN1, which is a typical secondary alcohol dehydrogenase that catalyzes the NAD-dependent stereospecific oxidation of (S)-1-phenylethanol to acetophenone in the degradation of ethylbenzene [58]. Both enzymes are classical members of the short-chain dehydrogenase/reductase (SDR) family with a reaction spectrum comprising the NAD[P][H]-dependent oxido-reduction of hydroxy/keto groups [59]. It is worth mentioning that strains containing psfG on no occasion also carry the hmfH gene (Fig. S3), suggesting that both enzymes could have analogous functions not previously considered, catalyzing HMFC conversion to FFCA (Fig. 1), which is the unidentified activity in strain ALS1267 that lacks a hmfH-encoding gene [21]. Nevertheless, entire hmf cluster, including hmfABCDEFG and psfAG genes, was cloned and transformed in strain KT2440, allowing it to grow on furfural as a sole carbon source but not on HMF, suggesting that an additional gene is still lacking to support growth on HMF [21], or not enough

Fig. 4. Presence and distribution of hmfFG, hmfH, psfG and psfA genes among proteobacterial genomes analyzed in this work. The presence of genes putatively encoding different enzymatic functions of the upper HMF pathway was analyzed in genomes of 99 proteobacterial species containing the hmfABCDEFG cluster responsible for the lower part of the route. The number of species harboring each gene is included in the circle representing it. External pie charts represent the distribution of each gene denoted in Venn diagram among different proteobacterial lineages. Blue, α-proteobacteria; orange, β-proteobacteria; green, γ-proteobacteria. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
expression of genes encoding the upper part of the route was achieved. A null mutant of psfC gene in ALS1267, or in related strains harboring a similar gene set, would help to clarify its role in HMF degradation pathway, which may lead to its use as a biocatalyst in the production of furan derivatives.

3.3. Furan utilization pattern of selected strains harboring different hmf/psf gene sets provides support for key role of some enzymes

Based on the analysis of presence/absence profiles of hmf/psf genes in proteobacterial genomes we selected representative strains harboring distinct gene sets to provide support for key role of some enzymes. Eight strains belonging to α- (1), β- (4), and γ- (3) Proteobacteria were selected to test their ability to grow on different intermediates of HMF pathway as sole carbon and energy source. The selected strains display clusterings ranging from full hmfABCDEF-psfA-psfG-psfH or hmfABCDEF-psfA-psfG gene sets to solely hmfABCDEF core genes. As shown in Fig. 5 the strains Cupriavidus basilensis DSM 11853 and Pseudomonas sp. ALS1279 harboring the hmfABCDEF-hmfFG-psfA-hmfH or hmfABCDEF-hmfFG-psfA-psfC genes respectively were able to grow on all intermediates tested with exception of FDCA, suggesting that both gene sets may provide a complete HMF pathway, and that psfG or hmfH genes could encode analogous functions. The inability to grow using FDCA as substrate is most probably related with the absence of a specific transporter system for this dicarboxylic acid. On the contrary, the strains Cupriavidus pinatubonensis JMP134 and Pseudomonas sp. ALS1131 harboring only hmfABCDEF genes are unable to grow in any intermediate except FA. In the case of Paraburkholderia xenovorans LB400 the presence of hmfFG genes along with hmfABCDEF could allow to this strain to proliferate on FDCA as sole carbon source, in addition to FA. Most probably this strain harbors a specific transport system for FDCA uptake that is not found in C. basilensis DSM 11853 or Pseudomonas sp. ALS1279, whose hmf clusters would be devoted to HMF consumption instead FDCA metabolization. Noteworthy, the presence of psfA in addition to hmfFG and hmfABCDEF genes in Pseudomonas mendocina VN230 and Azospirillum oryzae COC8 enable these strains to grow on FFCA besides FDCA and FA as substrates, supporting its role as a FFCA dehydrogenase. Finally, a remarkable case in HMF metabolism is represented by Paraburkholderia phytofirmans PsJN. It has been previously reported that this strain is able to grow on HMF although with very low growth rates, even 30 times slower than other strains harboring hmf gene clusters [21]. In the culture conditions tested by us, no significant growth in presence of HMF was assessed after 72 h of incubation. However, it was possible to measure significant growth in HMFCA and FFCA, supporting the notion that hmfH1 is able to supply the absence of psfA catalyzing consecutive oxidations of HMFCA to generate FDCA, but that the growth on HMF is limited by inefficient conversion into HMFCFA either by HmfH or housekeeping aldehyde dehydrogenases. Altogether the results of in vivo growth tests support the requirement of a hmfABCDEF-hmfFG-psfA-psfC-hmfH gene set and most probably a proficient aldehyde dehydrogenase acting on HMF for the complete metabolization of this substrate.

3.4. Transcriptional regulators associated to hmf genes are related to AraC- and LysR-type families.

On the contrary to the metabolic features of the HMF biodegradation route that has been roughly established, the molecular bases of transcriptional regulation of the pathway are mostly unknown. As previously reported, C. basilensis HMF14 carries hmf genes in two different gene clusters, both preceded by a LysR-type transcriptional regulator-encoding gene located in opposite orientation, termed hmfR1 and hmfR2, which could master the expression of HMF pathway by sensing substrates or intermediates of the route [20], but experimental evidence supporting this role has not been provided yet. In the case of P. putida Fu1, the inactivation of a homologous LysR-type regulatory protein (termed PsfB) abolished growth on

![Fig. 5. Growth on furan intermediates of several Proteobacterial species owing different hmf and psf gene sets.](image)
FA in this strain and their chemotactic responses towards furans [19,60]. Alternatively, it was reported the selection of an adaptive mutation in a regulatory gene coding for an activator of the AraC/XylS family in the furfural-degrading Pseudomonas pseudocaligenes CECT 5344 strain [61]. This point mutation affected the HTH region of the protein allowing efficient growth of the strain on FA and furfural; and on the contrary, the whole inactivation of the AraC-type regulator-encoding gene reversed the phenotype [61]. Both, LysR- and AraC-type regulators are commonly associated with carbon source assimilation, stress response and pathogenesis in bacteria [62,63]. Their involvement in transcriptional regulation of furan catabolism requires additional work to elucidate their specific roles. Accordingly, we found 82 bacterial genera from the total of selected strains harboring a homologous activator protein belonging to AraC-family in the context of hmf genes, whilst a transcriptional regulator belonging to LysR-family was identified in 49 species (Fig. S4). Notably, all bacterial strains that carry exclusively lower pathway-encoding genes (44) only own an AraC-type regulator (Fig. S4; Table S1), suggesting that hmfABCDE genes supporting FA metabolism are specifically mastered by this transcriptional regulator, comparable to the scenario reported by P. pseudoalcaligenes CECT 5344 [61]. Interestingly, the point mutation (L261R) that turns functional that AraC regulator is in fact a conserved residue (Arg) in the remaining homologues putatively involved in FA catabolism, as revealed by alignment of all AraC proteins identified in this study (data not shown), indicating that the inactive version of the regulator in wild type strain CECT 5344 is exceptional. On the other hand, all bacterial species that contain LysR-type regulators harbor upper and lower pathway genes (Fig. S4), indicating that this transcriptional regulator could be mainly associated to the presence of a full catabolic pathway for furans. As well, 33 bacterial species possessed both AraC- and LysR-type transcriptional regulators associated to hmf contexts (Fig. S4), analogous to the scenario described in clusters of P. putida Fu1 and ALS1267 [19,21]; suggesting that the upper pathway would be controlled by a LysR-type regulator and the lower route by the previously described AraC-type activator. The identity of the substrates/intermediates acting as effectors of transcriptional activation in strains harboring this putative regulatory architecture deserves further research. Finally, 11 selected strains mostly belonging to β-proteobacteria (7) harboring upper and lower pathway genes, contain solely a LysR-type regulator associated to hmf cluster without an accompanying AraC-type regulator (Table S1), therefore these genes could be regulated by a single activator/effect- tor pair. Nevertheless, 6 species belonging to Burkholderiaaceae family of β-Proteobacteria contain a second copy of LysR-type regulator, analogous to reported by strain HMF14 [20], suggesting that specifically in these strains this extra copy could be adapted to recognize an additional substrate/intermediate.

The identification of substrates/intermediates acting as effectors of the different transcriptional regulators putatively involved in furans catabolism is relevant not only to understand the molecular basis of pathway activation, but also from a biotechnological perspective on their potential for biosensing. Since cytotoxic, genotoxic, and carcinogenic effects of HMF and furans derivatives have been demonstrated in several animal models and mammalian cell cultures, its quantification becomes relevant in alimentary industry because this compound is produced during processing of several foodstuffs [64–67]. As an alternative to sophisticated chromatographic methods for HMF quantification, sensor protein-based biosensors would allow a fast and low-cost procedure to detect this compound and/or their transformation products. Transcriptional activators that sense HMF and/or intermediates of the catabolic route could be a valuable resource to construct biosensors able to detect these molecules, becoming highly desirable to describe the profile of effectors recognized by the proposed regulators.

3.5. Transporters-encoding genes associated to furans biodegradation

The role of transport in biodegradation of furans is largely unknown. The only piece of evidence about an active transport of furans comes from the establishment of a whole cell biocatalyst system in P. putida S12, in which the heterologous expression of a transporter termed HmfT1 had a considerable effect on the improvement in FDCA production from HMF mediated by HmFH activity of C. basilensis HMF14 [22]. This effect was attributed to a strong decrease in transient HMFCA accumulation, suggesting that the main function of hmfT1 gene product was uptake of this metabolite and therefore indicating the importance of transporter-encoding genes in furan metabolism [22,52]. HmfT1 belongs to Major Facilitator Superfamily, an ancient, extensive and ubiquitous family able to transport a broad spectrum of ions and solutes across membranes via facilitated diffusion, symport, or antiport [68]; and been assigned to the metabolite: H + symport subfamily [69]. We identified 29 strains that harbor HmfT1-transporter-encoding genes associated to hmf clusters, mostly belonging to α- (10) and β-proteobacteria (16) (Table S2). Most of these strains (24) possess upper pathway genes (Fig. S5), supporting that its main function is likely to be the transport of upper pathway substrates or intermediates as proposed in C. basilensis HMF14 [20,22]. In addition, we detected 39 species harboring a transporter-encoding gene associated to lower pathway genes (Table S2), which is homologous to a gene described in P. pseudoalcaligenes CECT 5344 and termed benE despite its linkage to hmf genes [61]. Similar to HmfT1, this transporter belongs to Major Facilitator Superfamily, although it was included in a different branch named Benzoate:H + symport subfamily, which first characterized member was the benzoate transporter encoded by benE from Acinetobacter calcoaceticus [69]. Interestingly, benE-like transporters were found primarily in α- (26) and γ-proteobacteria (11), and most strains (32) harbor only lower pathway genes (Fig. S5), suggesting that its main function could be FA transport.

Remarkably, we found that hmf clusters from 24 species belonging to α- (9) and β-proteobacteria (15) carry a putative transporter gene closely related to high affinity branched-chain amino acid transport system (LivKHMGF) from Escherichia coli (Fig. S5; Table S2), that relies in coupling of ATP-hydrolysis for uptake of substrates [70]; suggesting alternative mechanisms for furans transport. Remarkably, almost all strains (91) harbor one homologue of HmfT1-, BenE- or LivKHMGF-like transporters associated to hmf genes and its presence exclude the others (Fig. S5), except for Bordetella flabilis AU10664 that harbors two copies of hmfABCDEF cluster owing either HmfT1- or a LivKHMGF-like transporter associated to each copy. These observations suggest that these three different transporter types could accomplish analogous functions in furans transport. Finally, a diverse array of other transporter types could be also found associated to hmf genes but each to a much lower extent (Fig. S5), including homologues related to ABC-type nitrate/sulfonate/bicarbonate transport system (tauABC), TRAP-type C4-dicarboxylate transport system (dctPQM), TRAP-type uncharacterized transport system (Imp), tripartite tricarboxylate transporter substrate binding protein (tctC), and Na⁺:solute symporter family protein (SSS). This broad diversity of transporters putatively involved in furans catabolism would reflect that different substrates or intermediates in the route can be subject to active transport in degradative bacteria.

4. Conclusion

A systematic analysis of hmf genes in proteobacterial genomes has allowed establishment of their phylogenetic relationships, clusters organization and presence-absence patterns of catabolic
and accessory functions. From this study a previously unrecognized distribution of the HMF catabolic pathway could be observed, expanding the area of potential degradative microorganisms to be recruited in biodetoxifying processes to remove furan inhibitors from biomass hydrolysates. Most importantly, the persistent presence of some genes, or excluding patterns among them, allow us to predict the full gene set required to establish a functional HMF pathway in a different host. In addition to the FA-metabolizing lower route encoded by hmfABCD genes, the presence of hmfFG – psfA - hmfH/psfG genes it seems to be a key for proper HMF funneling into central metabolism. Furthermore, the recruitment of at least one of the main transporter types described as associated to hmf genes would be significant for a proficient metabolic engineering endeavor.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by FONDECYT 1201714, ANID PIA/Anillo ACT172128, ANID PIA/BASAL FB0002 grants.

Author contributions

RD performed most of bioinformatics studies, analyzed and interpreted all the achieved results, and wrote the original draft. FG-T performed preliminary bioinformatic analyses. DP-P conceived and designed throughout the studies, supervised the follow-up, and reviewed and edited the final manuscript. All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.017.

References

[1] Kawaguchi H, Hasumuma T, Ogino C, Kondo A. Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks. Curr Opin Biotechnol. 2016;42:30–9.

[2] Galkin KI, Ananikov VP. When Will 5-Hydroxymethylfurfural, the "Sleeping Giant" of Sustainable Chemistry, Awaken?. ChemSusChem. 2019;12(13):2976–82.

[3] Lalanne L, Nyanhongo GS, Guebitz GM, Pellis A. Biotechnological production and high potential of furan-based renewable monomers and polymers. Biotechnol Adv. 2019;48:107707. https://doi.org/10.1016/j.biotechadv.2021.107707.

[4] Hu L, He A, Liu X, Xia J, Xu J, Zhou S, et al. Biocatalytic Transformation of 5-Hydroxymethylfurfural into High-Value Derivatives: Recent Advances and Future Aspects. ACS Sustainable Chem Eng. 2018;6(12):19515–35.

[5] Global 5-hydroxymethylfurfural (5-HMF) (CAS 67-47-0) Market 2019 by Manufacturers, Regions, Type and Application, Forecast to 2024. Report ID: MSR1552060, Published Date: February 03, 2019.

[6] Rasmussen H, Sørensen HR, Meyer AS. Formation of degradation compounds via furfural and 5-(hydroxymethyl)furfural hydrolysates. Nat Chem Biol. 2010;6:1275–8.

[7] Dong H, Bao J. Metabolism: biofuel via biodetoxification. Nat Chem Biol. 2010;6:1275–8.

[8] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment. Nucleic Acids Res. 2002;30:3059–61.

[9] Parawira W, Tekere M. Biotechnological strategies to overcome inhibitors in lignocellulosic hydrolysates. Biotechnol Biofuels. 2015;8:198.

[10] Wang Y, Brown CA, Chen R. Industrial production, application, microbial biosynthesis and degradation of furanic compound, hydroxymethylfurfural (HMF), AIMS Microbiol 2018;4(2):261–73.

[11] Igeño MI, Sánchez-Clemente R, Población AC, Gujio MJ, Merchán F, Blasco R. Biodegradation of 5-(hydroxymethyl)-furfural and furan derivatives. Proceedings 2018;2(20):1283.

[12] Wierckx N, Koopman F, Ruijssenaars HJ, de Winde JH. Microbial degradation of furanic compounds: biochemistry, genetics, and impact. Appl Microbiol Biotechnol. 2011;92(6):1095–105.

[13] Nichols NN, Mertens JA. Identification and transcriptional profiling of Pseudomonas putida genes involved in furoic acid metabolism. FEMS Microbiol Lett. 2008;284(1):52–7.

[14] Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ. Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of Cupriavidus basilensis HMF14. Proc Natl Acad Sci U S A. 2010;107(11):4919–24.

[15] Crigler J, Esteman MA, Esteman E. Characterization of the Furfural and 5-Hydroxymethylfurfural (HMF) Metabolic Pathway in the Novel Isolate Pseudomonas putida AL51267. Appl Biochem Biotechnol. 2020;190(3):918–30.

[16] Wierckx N, Elink Schuurman TD, Blank LM, Ruijssenaars HJ (2015) Whole-Cell Biocatalytic Production of 2,5-Furandicarboxylic Acid. In: Kamn B. (eds) Microorganisms in Biorefineries. Microbiology Monographs, vol 26. Springer, Berlin, Heidelberg.

[17] Hossain GS, Yuan H, Li J, Shin H-D, Wang M, Du G, et al. Metabolic engineering of Raoultella ornithinolytica DSM 15950 for production of 2,5-furandicarboxylic acid from 5-hydroxymethylfurfural. Appl Environ Microbiol. 2017;83(4):1198–201. https://doi.org/10.1128/AEM.02112-16.

[18] Guarnieri MT, Ann Franden M, Johnson CW, Beckham GT. Conversion and assimilation of furfural and 5-(hydroxymethyl)furfural by Pseudomonas putida KT2440. Metab Eng Commun. 2017;4:22–8.

[19] Baptista SL, Costa CE, Cunha JT, Soares PO, Domingues L. Metabolic engineering of Saccharomyces cerevisiae for the production of top value chemicals from biomass-derived carbohydrates. Biotechnol Adv. 2021;47:107679. https://doi.org/10.1016/j.biotechadv.2021.107677.

[20] Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ. Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid. Biotechnol Biofuels. 2010;3(1):6291–6.

[21] Herrgärd M, Panagiotou G. Analyzing the genomic variation of microbial cell factories in the era of "New Biotechnology". Comput Struct Biotechnol J. 2013;4(3):e20120012. https://doi.org/10.5936/csbi.201212012.

[22] Johnson M, Zaretsky I, Ravelo V, Mereshuk Y, McGlinnis S, Madden TL. NCBI BLAST: a better web interface. Nuc Acids Res. 2008;36(Web Server issue): W5–9.

[23] Chen IA, Chu K, Palamiappan K, Pillay A, Huang J, Huntmamnen M, Varghese N, White JR, Seibelhi R, Smirnova T, Kirton S, Jungbluth SP, Woyte Y, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC, IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. NuclAcids Res. 2019; 47(D1):D666–D677.

[24] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucl Acids Res. 2018;46(18):D8–D13.

[25] Lu S, Wang J, Chitfas T, Derbystyke MK, Geer RC, Gonzales NR, Czudak M, Hong J, Shin H-D, Wang M, et al. Metabolic engineering of Sphingomonas sp. C2-2507 for production of 2,5-furandicarboxylic acid from 5-hydroxymethylfurfural. Appl Microbiol Biotechnol. 2017;11(11):4919–24.

[26] Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.

[27] Katoz K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160–6.

[28] Letunic I, Bork P. Interactive Tree Of Life (ITOL) v4: recent updates and new developments. Nucl Acids Res. 2019;47(W1):W256–5.
Pham NN, Chen C-Y, Li H, Nguyen MTT, Nguyen PKP, Tsai S-L, et al. Engineering
Marshall SA, Payne KAP, Leys D. The UbiX-UbiD system: The biosynthesis and
Donoso R, González-Toro F, Guajardo-Parra M, Araya-Nail M, Farkas C,
Kniemeyer O, Heider J. (...
Saaret A, Balaikaite A, Leys D. Biochemistry of prenylated-FMN enzymes.
Igeño MI, Macias D, Blasco R. A Case of Adaptive Laboratory Evolution (ALE):
Maddocks SE, Oyston PCF. Structure and function of the LysR-type
Wu S, Liu Q, Tan H, Zhang F, Yin H. A Novel 2,5-Furandicarboxylic Acid Biosynthesis
Hsu C-T, Kuo Y-C, Liu Y-C, Tsai S-L. Green conversion of 5-
Payne KAP, Marshall SA, Fisher K, Cliff MJ, Cannas DM, Yan C, et al. Enzymatic
Viñambres M, Espada M, Martínez AT, Serrano A. Screening and Evaluation of
Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. Arac/XylS family of
Dijkman WP, Fraaije MW. Discovery and characterization of a 5-
Yuan H, Li J, Shin H-D, Du H, You J, Li H, Ke C, Zhang X, Yu J, Zhao T. Luteimonas
Pao SS, Paulsen IT, Saier MH. Major facilitator superfamily. Microbiol Mol Biol
Adams MD, Wagner LM, Graddis TJ, Landick R, Antonucci TK, Gibson AL, et al.
Lin J, Yang G, Tang J, Li Z, Yu Z, Zhang Li. Pseudoanthomonas composti sp. nov.,
Wu G, Liu Y, Li Q, Du H, You J, Li H, Ke C, Zhang X, Yu J, Zhao T. Luteimonas
Koszucka A, Nowak A. Thermal processing food-related toxicants: a review.
Islam MN, Khalil MI, Islam MA, Gan SH. Toxic compounds in honey. J Appl
Nichols NN, Lunde TA, Graden KC, Hallock KA, Kowalchyk CK, Southern RM,
Raúl A. Donoso, Fabián González-Toro and D. Pérez-Pantoja Computational and Structural Biotechnology Journal 19 (2021) 2160–2169