SusC/D-like proteins in *Gammaproteobacteria* that utilize fructans

Marie-Katherin Zühlke[1,2*], Alexandra Bahr[1,3], Daniel Bartosik[1,2], Vipul Solanki[4], Michelle Teune[5], Frank Unfried[1], Tristan Barbeyron[6], Elizabeth Ficko-Blean[6], Lionel Cladière[6], Alexandra Jeudy[6], Anne Susemihl[5,7], Fabian Hartmann[1], Diane Jouanneau[6], Murielle Jam[6], Matthias Höhne[5,8], Mihaela Delcea[5], Uwe T. Bornscheuer[5], Dörte Becher[9], Jan-Hendrik Hehemann[2,4,10], Mirjam Czjzek[6], Thomas Schweder[1,2*]

[1] Institute of Pharmacy, University of Greifswald, Germany

[2] Institute of Marine Biotechnology, Greifswald, Germany

[3] Helmholtz Institute for One Health, Helmholtz Centre for Infection Research HZI, Greifswald, Germany

[4] Max Planck Institute for Marine Microbiology, Bremen, Germany

[5] Institute of Biochemistry, University of Greifswald, Germany

[6] Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Sorbonne Université, CNRS, Brittany, France

[7] Internal Medicine C, University Medicine Greifswald, Germany

[8] Institute of Chemistry, Technical University of Berlin, Germany

[9] Microbial Proteomics, Institute of Microbiology, University of Greifswald, Germany

[10]Center for Marine Environmental Sciences, University of Bremen

[*]corresponding authors

Marie-Katherin Zühlke, e-mail: marie-katherin.zuehlke@uni-greifswald.de

Thomas Schweder, e-mail: schweder@uni-greifswald.de

Biological Sciences/Biochemistry

Key words

SusD-like proteins / fructan / inulin / glycoside hydrolase family 32 / *Gammaproteobacteria*

Abstract

SusC/D-like proteins are essential components of glycan utilization machineries in *Bacteroidota*, but remain unknown in other bacterial phyla. The glycan-binding SusD-like protein forms a 'lid' on top of the SusC-like TonB-dependent transporter (TBDT) and both are structurally designed to function as a complex in sugar uptake. In comparison, *Gammaproteobacteria* import glycans using 'classical' TBDTs without an accessory SusD-like protein. We have now identified a SusD-like protein and a SusC-like TBDT in a fructan polysaccharide utilization locus (PUL) of the marine gammaproteobacterium *Pseudoalteromonas distincta*, which are tandemly organized as in *Bacteroidota*. Proteome analysis revealed an increased production of PUL-encoded proteins during growth on inulin- and levan-type fructans. However, *P. distincta* preferred inulin over plant-derived levan and hardly grew on bacterial levan. Further analysis showed that the SusD-like protein has a weak affinity (K_a 43 M^{-1}) for oligosaccharides from inulin. The PUL-encoded glycoside hydrolase from family 32 (GH32) hydrolyzes inulin and plant-derived levan, but also has a low activity on bacterial levan, which confirms the growth experiments. Comparative genomics identified further SusC/D-like proteins in *Gammaproteobacteria* genomes, most of which (83%) were encoded in fructan PULs.

Significance statement

It is a mystery why the Gram-negative *Bacteroidota* and *Gammaproteobacteria* use different transport systems for glycan utilization. In contrast to this observation, we found that the gammaproteobacterium *Pseudoalteromonas distincta* uses a *Bacteroidota*-like transport complex in fructan uptake, composed of a glycan-binding SusD-like protein and a SusC-like TonB-dependent transporter. Proteome analysis revealed that the SusC/D-like proteins and a fructan-degrading enzyme are produced at high levels in fructan-grown cells. Weak affinity of the SusD-like protein to inulin oligosaccharides and enzyme activity on fructans confirmed their

fructan specificity. Further SusC/D-like pairs were detected in other *Gammaproteobacteria*, mostly associated with fructan-related gene clusters. This study thus describes marine fructan use by *Gammaproteobacteria* and indicates an exchange of protein functions between phyla.

Introduction

Fructans are polyfructose chains, in which fructosyl units are β(2,1)- or β(2,6)-linked to sucrose, referred to as inulin or levan, respectively. The production of fructans permeates all kingdoms of life, except animals. Plants and bacteria are known producers of fructans, while reports on fungi and especially on algae are rare (1). While fructans serve as storage compounds or osmolytes in plants, they are frequently found as exopolysaccharides (EPS) in Gram-positive and Gram-negative bacteria. Bacteria cover their surface with these protective glycans that are essential components in host-pathogen interaction, increase bacteriophage evasion and antibiotic resistance (2). Bacterial fructan production includes species from the genera *Erwinia*, *Bacillus*, *Pseudomonas* or *Zymomonas*, but also lactic acid bacteria (3). Consequently, the ability to degrade fructans is widespread among terrestrial bacteria (4) and was investigated in *Bacteroidota* of the human intestine (5-8). In comparison, the occurrence and the relevance of fructans in marine ecosystems is underexplored, with respect to their production and their degradation.

Bacteroidota and *Gammaproteobacteria* are important players in marine glycan degradation. In the *Bacteroidota* phylum, gene clusters encoding glycan-disassembling carbohydrate-active enzymes (CAZymes) and glycan-transporting SusC/D-like complexes were defined as polysaccharide utilization loci (PULs), where ‘sus’ refers to homologs of the starch utilization system (sus) of *Bacteroides thetaiotaomicron* (9, 10). The glycan-binding SusD-like protein, which is tethered to the outer membrane, transmits the oligosaccharides released by extracellular CAZymes to the SusC-like TonB-dependent receptors or transducers (referred herein as transporters, TBDTs), the so-called ‘pedal-bin’ mechanism (11). Compared to ‘classic’ TBDTs, SusC-like TBDTs are increased in size to accommodate the SusD-like protein (12). In comparison, carbohydrate utilization clusters of *Gammaproteobacteria* (phylum

Pseudomonadota), often referred to as PULs as well, lack SusD-like proteins and the "SusC-like" specific characteristics of TBDTs.

PULs targeting fructans are considered "single-family PULs" (13), which means that they only encode one CAZyme function, in this case the glycoside hydrolase (GH) family 32, GH32. GH32s include, e.g., endo- and exo-inulinases/levanases or invertases. Although inulinase activity of marine fungi and bacteria (14-17) or the presence of GH32-containing PULs in metagenome-assembled genomes of marine bacteria (18, 19) have been reported, in-depth biochemical and structural analyses of the corresponding CAZymes, as in the case of the *Thermotoga maritima* GH32 (17, 20), are rare. In addition, there are only a few reports on marine bacteria that produce fructans/fructose-containing EPS (21, 22) or that encode the fructan-synthesizing GH68 and fructan-recycling GH32 (23).

We previously investigated the PUL repertoire of *Pseudoalteromonas distincta* ANT/505, which produces vesicles and vesicle chains from its outer membrane (24). The PUL analysis already identified a GH32-encoding PUL. Surprisingly, we now discovered that this fructan PUL also encoded a SusC-like TBDT/SusD-like protein pair, which was thought only to be found in *Bacteroidota*. Therefore, the proteins were probably acquired from a *Bacteroidota* ancestor and are described here for the first time for *Gammaproteobacteria*.

In this study, we thus investigated the utilization of fructans by the marine gammaproteobacterium *P. distincta* using proteomics and by characterizing the PUL-encoded SusD-like protein and GH32. We further explore the distribution of *Gammaproteobacteria* utilizing the SusC/D-like proteins in the marine environment and beyond.

Results

Pseudoalteromonas distincta *degrades fructose-containing substrates*

Growth studies using fructans of terrestrial origin demonstrated that *P. distincta* is adapted to a fructose-based diet. *P. distincta* grew to similar optical densities (OD_{600nm} 2-2.5) within one day using inulin, oligosaccharides from inulin (FOS_{INU}), sucrose or fructose (Fig. 1A). In comparison, *P. distincta* reached low optical densities (OD_{600nm} 0.4-0.5) using plant-derived

levan and hardly grew on bacterial levan (Fig. 1A, see also control without substrate Fig. S1A). Bacterial levan has a high molecular weight (>1000 kDa (25)) and is branched (26) compared to plant-derived levan (~12.5 kDa) or small-sized inulin (degree of polymerization, DP, 2-60).

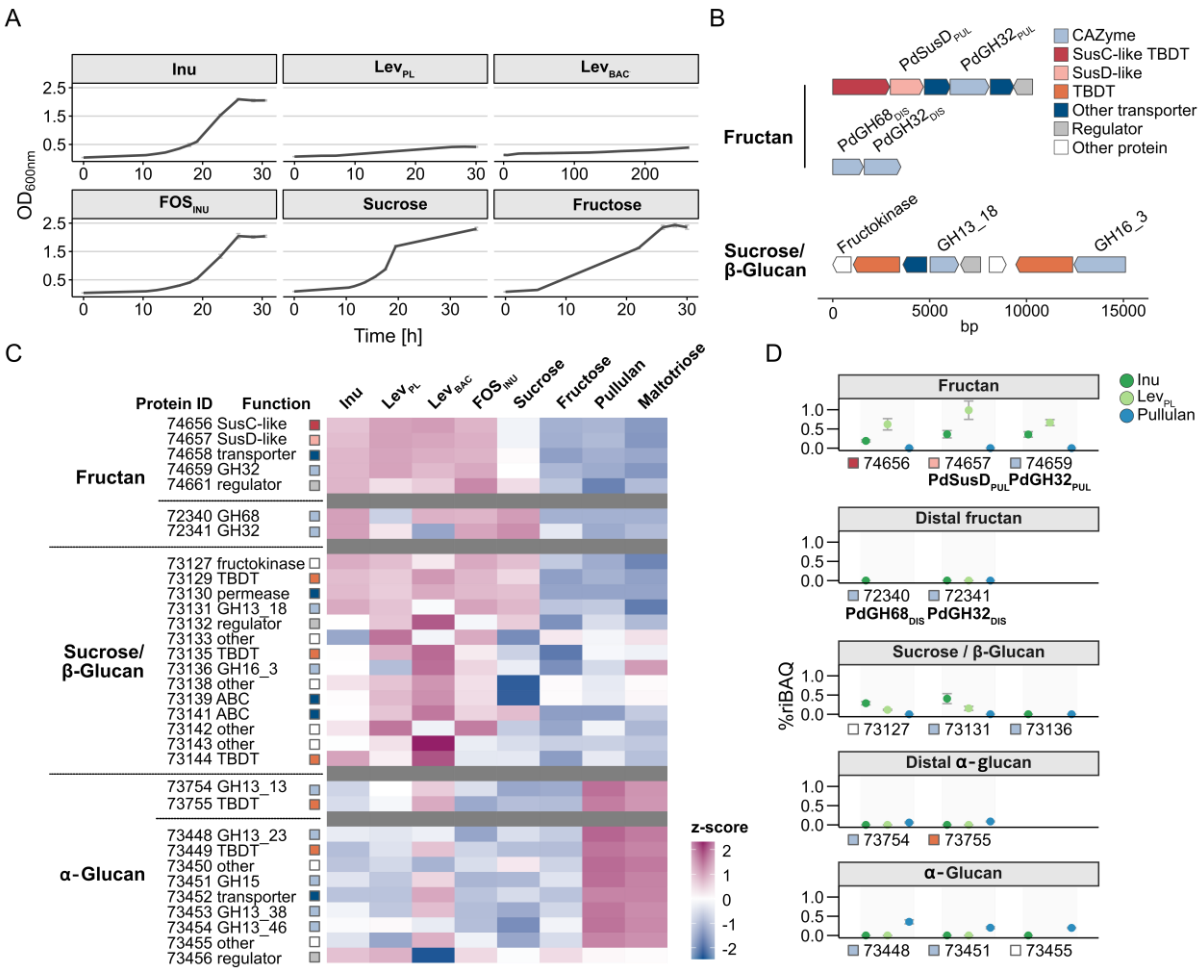

Fig. 1. The gammaproteobacterium *P. distincta* grows on fructans, which is presumably enabled by a PUL encoding a GH32 and SusC/D-like proteins. (A) *P. distincta* preferred inulin and inulin-derived oligosaccharides (FOS_{INU}) over levans. Due to slow growth, there is a different time scale for bacterial levan. Controls without the addition of a carbon source are shown in Fig. S1A. (B) The genomic context of two GH32s, encoded in a fructan PUL ($PdGH32_{PUL}$) and distal to the PUL ($PdGH32_{DIS}$), are shown. The CAZyme repertoire of a genomic region likely specific for sucrose (GH13_18) and β-glucans (GH16_3) is also shown. The cluster might be even longer, see Fig. 1C. (C) Differential protein production in *P. distincta*. Relative protein abundance values (%riBAQ, relative intensity-based absolute quantification, giving the relative protein abundance within the sample, n=3) were log-normalized and z-scored. (D) Means of %riBAQ values and corresponding standard deviation were plotted for selected proteins (for individual values and statistics, please refer to Supplementary Dataset 1). Protein accessions were abbreviated, e.g., 74657 refers to EGI74657.1. Inu: Inulin, Lev_{PL}: plant-derived levan, Lev_{BAC}: Bacterial levan. FOS_{INU} fructo-oligosaccharides from inulin.

While all other *P. distincta* PULs correspond to classical *Gammaproteobacteria* PULs, which lack the SusC-like specification of TBDTs and a SusD-like protein (24), the GH32-encoding cluster classifies as a *Bacteroidota*-like PUL (Fig. 1B). Upstream of the PUL-encoded GH32

(PdGH32$_{PUL}$, EGI74659), two genes were annotated as “putative outer membrane protein, probably involved in nutrient binding”. Unexpectedly, database searches identified them as a SusD-like protein (PdSusD$_{PUL}$, EGI74657) and a SusC-like TBDT (EGI74656) (Fig. 1B). In addition to the receptor/β-barrel domain PF00593 and plug domain PF07715 for TBDTs, the SusC/RagA model TIGR04056 (SusC *B. thetaiotaomicron*, RagA *Porphyromonas gingivalis*) for the SusC-like subclade was only identified in the case of the fructan PUL-encoded TBDT (EGI74656). In line with this, the Pfam family for SusD/RagB homologs PF07980 was identified for PdSusD$_{PUL}$. To the best of our knowledge, this is the first report of SusC/D-like proteins in *Gammaproteobacteria*.

Distal to the fructan PUL, an additional *gh32* gene (distal PdGH32$_{DIS}$, EGI72341) is co-localized with a *gh68* (distal PdGH68$_{DIS}$, EGI72340), suggesting fructan-production (PdGH68$_{DIS}$) and -recycling (PdGH32$_{DIS}$) (Fig. 1B). Another *Gammaproteobacteria* PUL encodes proteins that presumably cleave sucrose (GH13 subfamily 18, GH13_18, EGI73131) and metabolize fructose (fructokinase, EG73127) (Fig. 1B). However, the PUL also encodes a GH16_3 (EGI73136), which could indicate β-glucan use. We therefore refer to the PUL as sucrose/β-glucan genetic region, as it is unknown whether this genetic region is actually one PUL dedicated to a common substrate or whether it is two PULs.

Despite better growth on inulin, levan also stimulates protein production of fructan PUL proteins

Next, we analyzed whole-cell proteomes of *P. distincta* grown on fructose-containing substrates, listed in Fig. 1A, and using the α-glucan pullulan and its building block maltotriose as controls (Fig. S1A and B). Inulin, FOS$_{INU}$ and plant-derived levan, but also the bacterial levan, induced production of fructan PUL-encoded proteins (Fig. 1C), which were highly abundant in the samples (Fig. 1D, Fig. S2, Supplementary Dataset 1). In addition, fructans triggered a response of the sucrose/β-glucan genetic region (Fig. 1C). However, high protein abundance was only detected for the sucrose-related part of the PUL (Fig. 1D, Fig. S3). Therefore, the *gh16_3* and the co-localized *tbdt* may be a PUL separate from the sucrose PUL. Compared to proteins encoded in the fructan and sucrose PUL, protein abundance of the distal

$PdGH68_{DIS}$ and $PdGH32_{DIS}$ was very low and $PdGH68_{DIS}$ was not quantified with α-glucans, plant-derived levan and fructose (Fig. 1C, Fig. S4).

Unexpectedly, protein production of α-glucan-related proteins was increased with growth on bacterial levan (Fig. 1C), although protein abundance was much lower compared to pullulan or maltotriose samples (Fig. S5). Since *P. distincta* hardly grew on bacterial levan, this reduced growth could induce cell stress and lysis. Therefore, the increased protein abundance of α-glucan-related proteins might indicate that *P. distincta* uses α-glucan-type storage polysaccharides, like glycogen, during cultivation with bacterial levan. Apart from using its own storage glycans, *P. distincta* might also use α-glucans released from bacterial necromass as recently described (27). However, *P. distincta* contains distinct protein-coding genes for glycogen synthesis and recycling, which were produced across substrates (EGI74393-98, EGI74399 was not quantified, Supplementary Dataset 1).

$PdSusD_{PUL}$ binds FOS_{INU} with weak affinity

While growth experiments showed a preference for inulin, both inulin and levan stimulated the production of fructan PUL-encoded proteins. Therefore, we further investigated the proteins that are presumably involved in fructan use, which included $PdSusD_{PUL}$. A distance matrix alignment analysis (DALI) (28) using the AlphaFold2 (AF)-predicted structure of $PdSusD_{PUL}$ (Fig. 2A) identified the levan-binding SusD-like protein (BT_1762) from *B. thetaiotaomicron* VPI-5482^{T} as the closest homolog (PDB IDs: 5LX8 BT_1762; 5T4Y and 5T3R BT_1762/63, SusC/D) with z-scores >50. Both sequences share 59% sequence similarity and a core RMSD of 1.17 Å, which was determined using SSM superimpose in COOT (29). Nevertheless, sequence similarity of $PdSusD_{PUL}$ to inulin-binding SusD-like proteins from *Bacteroidota* of the human intestine is similarly high (54-56%, 57-62 % levan-binding proteins, Fig. S6). In addition, some residues known to be involved in levan-binding of the SusD-like protein (BT_1762) from *B. thetaiotaomicron* VPI-5482^{T} are missing in $PdSusD_{PUL}$ and inulin-binding proteins (Fig. S6). In line with this, in a large-scale phylogenetic analysis (1,055 SusD-like protein sequences),

PdSusD$_{PUL}$ clusters neither with levan- nor with inulin-specific SusD-like proteins from *Bacteroidota* of the human intestine (Fig. S7).

Affinity gel electrophoresis (1% of glycan applied) and surface plasmon resonance did not detect binding of PdSusD$_{PUL}$ to inulin or levan (data not shown). Similarly, isothermal titration calorimetry (ITC) was not able to detect binding to inulin, levan, sucrose or fructose at the concentrations used (data not shown, see Materials and Methods). However, ITC revealed weak affinity (K_a 43 M^{-1}) of PdSusD$_{PUL}$ to FOS$_{INU}$ (Fig. 2B, Fig. S8A). The interaction was enthalpically favorable and entropically unfavorable (Fig. S8B), as already observed for the levan-binding SusD-like protein (BT_1762) from *B. thetaiotaomicron* VPI-5482^T (5). Attempts to obtain a 3D crystal structure for PdSusD$_{PUL}$ to gain further insight into ligand binding were not successful.

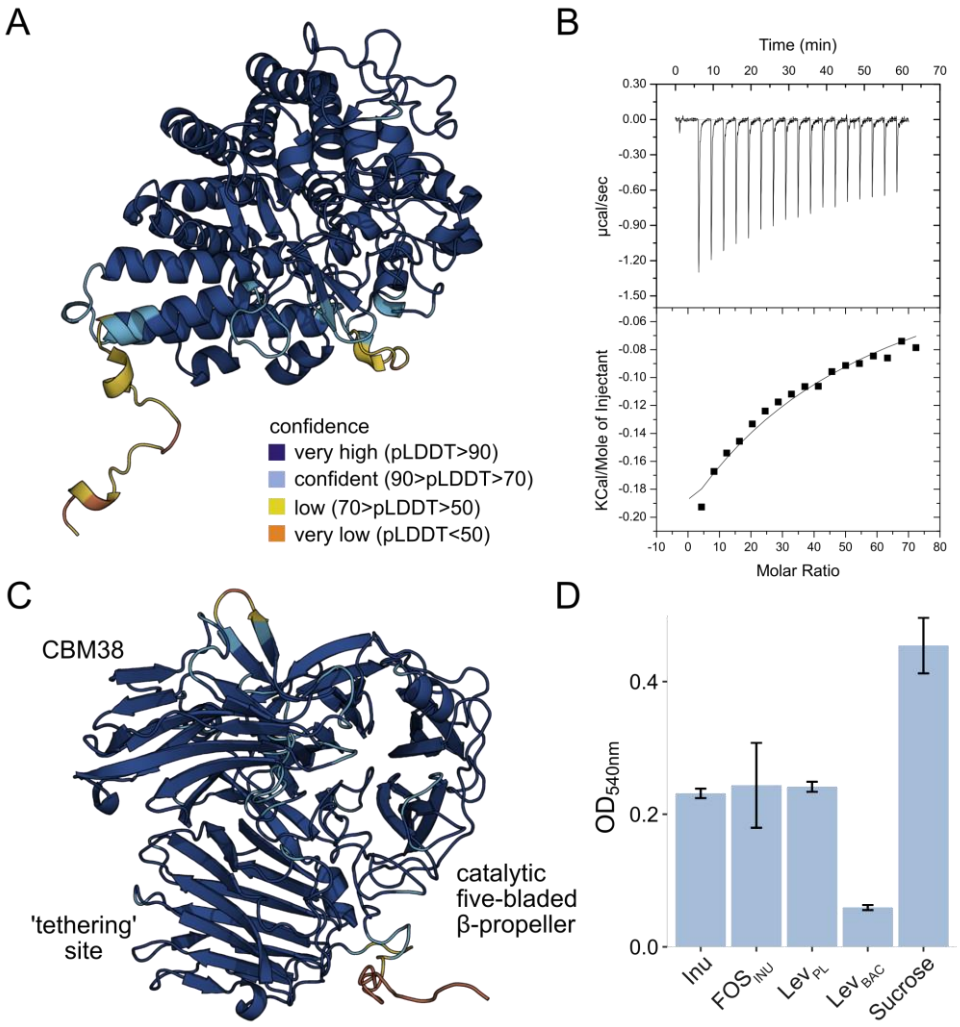

Fig. 2. The PUL-encoded SusD-like protein weakly binds FOS$_{INU}$, while the PUL-encoded GH32 cleaves inulin and levan. (A) AF-predicted structure of PdSusD$_{PUL}$. (B) Binding of PdSusD$_{PUL}$ to inulin-

derived oligosaccharides FOS_{INU}. Raw titration data (upper panel) and integrated heat peaks (lower panel) as a function of the molar ratio of ligand. (C) AF-predicted structure of $PdGH32_{PUL}$. (D) Activity of $PdGH32_{PUL}$ on different fructans and sucrose as detected by RSA, shown as mean with standard deviation (n=3). Values were corrected against controls without CAZyme. As a reference, see also Fig. S12A (values non-active distal $PdGH32_{DIS}$). The color code of 3D protein structures in (A) and (C) refers to the pLDDT confidence scores. Predicted signal peptides were omitted. Inu: Inulin, FOS_{INU} fructo-oligosaccharides from inulin, Lev_{PL}: Plant-derived levan, Lev_{BAC}: Bacterial levan.

The SusC/D complex in Pseudoalteromonas distincta

In addition to the high structural similarity of the AF-predicted 3D structure of $PdSusD_{PUL}$ with the 3D crystal structure of the levan-binding SusD-like protein (BT_1762), specific loops that enable interaction between the SusD and SusC (BT_1762/63) in *B. thetaiotaomicron* VPI-5482^T (11, 30) are also present in the *P. distincta* proteins using AF predictions (Fig. S9). Consistent with the above-mentioned sequence similarity of the inulin- and levan-associated SusD-like proteins, the similarity of their co-encoded SusC-like proteins to the *P. distincta* SusC-like protein was 57–63% (coverage >90%). The coverage was lower compared to the SusD-like protein (≥99%) because the N-terminal carboxypeptidase D regulatory-like domain (PF13715), frequently found in *Bacteroidota* SusC-like proteins (31), was not present in the *P. distincta* SusC-like protein. In *B. thetaiotaomicron* VPI-5482^T it was hypothesized that this domain provides specificity for inner membrane TonB proteins (30).

The fructan PUL-encoded $PdGH32_{PUL}$ efficiently hydrolyzes inulin, but also plant-derived levan

In general, GH32s may be endo- or exo-lytic. In endo-GH32s, a wider catalytic groove accommodates the fructan chain (6, 32). Attempts to obtain 3D crystal structures of $PdGH32_{PUL}$ were not successful. Using the AF-predicted structure of $PdGH32_{PUL}$ (Fig. 2C), a DALI search (28) revealed highest similarities to the exo-inulinase of *Aspergillus awamori* (z-score >48, PDB IDs 1Y4W, 1Y9G, 1Y9M (33)). The catalytic- and intermediate-stabilizing residues, the catalytic triad, of $PdGH32_{PUL}$ superimpose with those of the *A. awamori* protein, but also with those of other characterized GH32s of endo- or exo-activity (Fig. S10). However, AF predictions revealed a ‘funnel’-like binding pocket of $PdGH32_{PUL}$ as in exo-GH32s (Fig. S11A). This was also confirmed by comparing groove-delimiting loops and residues (Fig. S11B), which

correspond to the exo-cleaving GH32 of *A. awamori* (33). Here, the -3 subsite that enables endo-cleavage in GH32 from *Aspergillus ficuum* (32) is blocked by the presence or orientation of tryptophan residues (Fig. S11B).

$PdGH32_{PUL}$ hydrolyzed inulin, FOS_{INU} and plant-derived levan, but also exhibited low activity on bacterial levan (Fig. 2D, Fig. S12A-D), which is presumably more challenging due to its size and branching. Corroborating the AF-predicted structure, FACE analysis indicated exo-activity of $PdGH32_{PUL}$ as sucrose was still hydrolyzed (Fig. S12A and D). These findings are in line with results from *B. thetaiotaomicron*, where exo-GH32s from *B. thetaiotaomicron* degrade inulin and levan, whereas the activity of the endo-GH32 is restricted to levan (5).

A special feature of the PUL-encoded $PdGH32_{PUL}$ compared to the distal $PdGH32_{DIS}$ is an additional CBM38, which protrudes between blade IV and V of the five-bladed β-propeller catalytic module (Fig. 2C, Fig. S13). CBM38 is known to bind inulin and levan (34).

Pseudoalteromonas distincta *presumably produces its own fructan*

Inulo- and levansucrases belong to the family GH68. Sucrases hydrolyze sucrose and add the released fructosyl unit to sucrose or other acceptors. The non-PUL-encoded distal $PdGH68_{DIS}$ and $PdGH32_{DIS}$ thus suggest that *P. distincta* can synthesize fructans ($PdGH68_{DIS}$), presumably as EPS, and recycle it ($PdGH32_{DIS}$). However, under the conditions used, we only detected high sucrose-hydrolyzing activity for $PdGH68_{DIS}$ (Fig. S12) and no production of higher molecular weight compounds. In addition, we could not capture any activity for $PdGH32_{DIS}$ (Fig. S12). While the PUL-encoded $PdGH32_{PUL}$ was predicted to have a 'funnel'-like catalytic groove, which suggests exo-activity, the AF-predicted binding-groove of the distal $PdGH32_{DIS}$ is larger (Fig. S11A). Correspondingly, closest homologs of DALI searches using $PdGH32_{DIS}$ were a levan fructotransferase from *Paenarthrobacter ureafaciens* K-2032 (formerly *Arthrobacter ureafaciens*, LFTase, PDB IDs 4FFF, 4FFG, 4FFH, 4FFI (35)) and the endo-inulinase from *A. ficuum* (3SC7, 3RWK (32)) with z-scores >46. The *P. ureafaciens* LFTase releases cyclic disaccharides (DFA-IV) from inulin for which a wider groove is provided

(35). Nevertheless, specific residues in the loop, which form the pocket to release cyclic disaccharides in LFTases (4), differ in $PdGH32_{DIS}$ (Fig. S14). In addition, similar to the PUL-encoded $PdGH32_{PUL}$, the -3 subsite, present in the endo-active GH32 from *A. ficuum* is blocked in the distal $PdGH32_{DIS}$ (Fig. S11B), again suggesting exo-activity.

Intracellular or extracellular fructan-disassembly?

The binding of FOS_{INU} by $PdSusD_{PUL}$ and the exo-activity of $PdGH32_{PUL}$ raised the question of the cellular localization of the CAZymes. Therefore, we analyzed the outer membrane-, the outer membrane vesicle-, the extracellular- and the intracellular soluble proteome of inulin- and plant levan-grown *P. distincta*. In agreement with whole-cell proteomes, fructan PUL-encoded proteins were highly abundant in all analyzed fractions including the intracellular soluble proteome (Fig. 3, Supplementary Dataset 2), which emphasizes their elevated production. While PSORTb and DeepLocPro predict a cytoplasmic or extracellular localization of $PdGH32_{PUL}$, respectively, the CAZyme was abundant in all subcellular proteomes, including the outer membranes and outer membrane vesicles (Fig. 3, Supplementary Dataset 2). In combination with the presence of a lipoprotein signal peptide, this indicates an outer membrane attachment. In comparison, the distal $PdGH68_{DIS}$ was the most abundant protein in the extracellular proteome of inulin-grown cells, but was not abundant in the other fractions (Fig. 3, Supplementary Dataset 2), which is in agreement with whole-cell proteomes (Fig. 1D). Extracellular production of fructans is known for Gram-negative bacteria (36). The distal $PdGH32_{DIS}$ was exclusive to the extracellular fraction (Fig. 3, Supplementary Dataset 2).

In addition to the fructan PUL-encoded SusC/D-like proteins, the TBDT encoded in the sucrose/β-glucan PUL (sucrose part) was also present in high quantities in most fractions, while the GH13_18 and fructokinase were enriched in the intracellular soluble proteome (Fig. 3, Supplementary Dataset 2). Fructans are synthesized from sucrose, therefore the presence of fructans could signal the availability of sucrose and increase the production of the proteins encoded in both PULs. Moreover, the fructokinase might phosphorylate fructose released by GH32 activity.

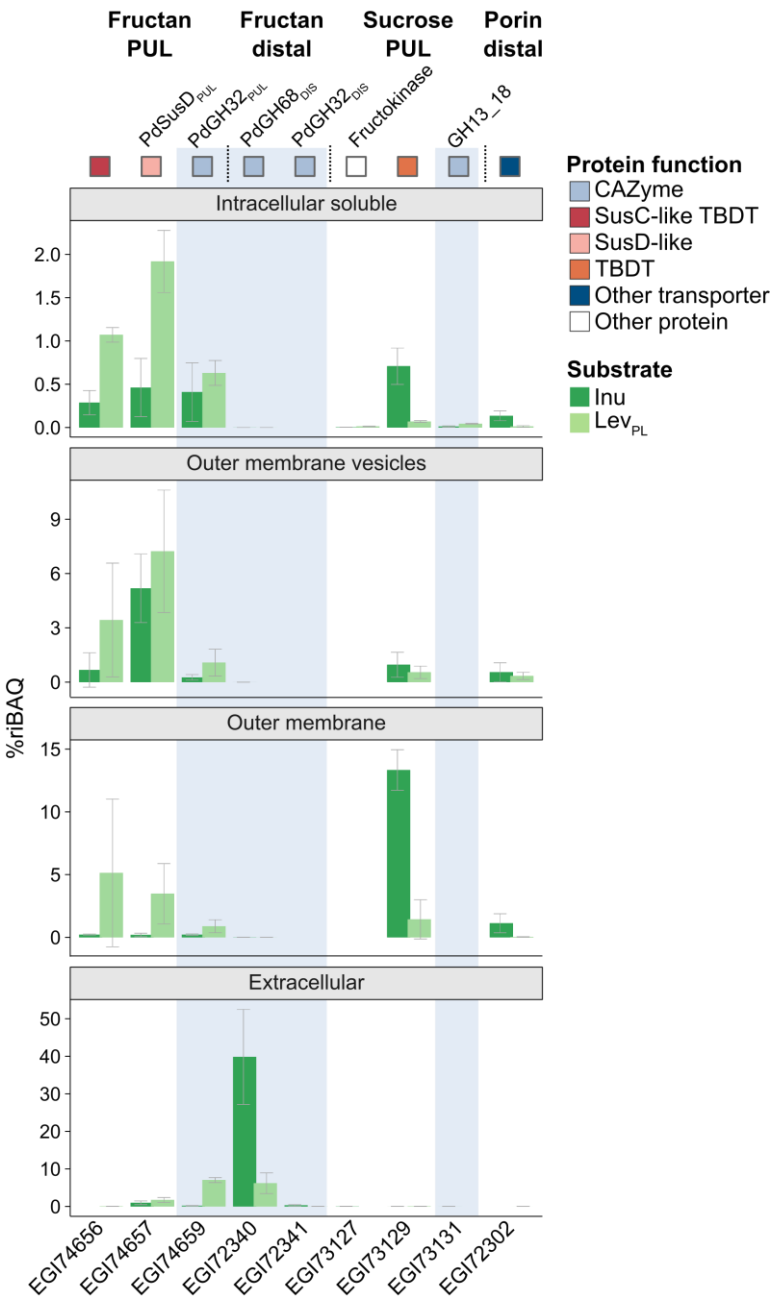

Fig. 3. The PUL-encoded PdGH32$_{PUL}$ is highly abundant in each fraction, while the distal PdGH68$_{DIS}$/PdGH32$_{DIS}$ are associated with the extracellular proteome. Relative abundance (%riBAQ) of selected proteins encoded in the fructan and sucrose PUL as well as of the distal PdGH68$_{DIS}$/PdGH32$_{DIS}$ and a porin. Protein abundance refers to means (n=3) with corresponding standard deviation, for individual values please refer to Supplementary Dataset 2. Protein abundance should only be compared within the same subcellular fraction. Inu: Inulin, Lev$_{PL}$: Plant-derived levan.

SusC/D-like proteins in Gammaproteobacteria*: only in fructan PULs?*

The phylogenetic analysis identified 13 additional gammaproteobacterial SusD-like protein sequences within a huge number of *Bacteroidota* sequences (1,005; in a total of 1,055 protein

sequences; Fig. S7). All gammaproteobacterial sequences had a GH32 context (data not shown). To further investigate this, we searched the GEM and UGCMP databases for gammaproteobacterial SusD-like sequences. In a total of 4,681 representative *Gammaproteobacteria* genomes, 32 clusters (31 genomes) encoded SusD-like proteins, out of which 26 were of marine origin (Fig. 4). Most SusD-like protein-coding sequences occurred as SusC/D-like pairs (91%) and were co-encoded with GH32s (75%) indicative of PULs (Fig. 4). The number might be even higher due to incompleteness of clusters (occasionally fragmented genomes). Noteworthy, SusD-like sequences were never encoded upstream of 'classical' TBDTs. In most cases, the presence of a SusC/D-containing fructan PUL did not coincide with a distal GH68/GH32 pair.

Altogether, 631 GH32s were identified in 574 gammaproteobacterial clusters (426 genomes) using this dataset, of which 89% occurred alone, indicating that many GH32s are not PUL-encoded. 4% of the identified GH32s relate to the above-mentioned SusC/D-containing PULs, whereas 7% co-occurred with TBDTs instead, indicating classical *Gammaproteobacteria* PULs.

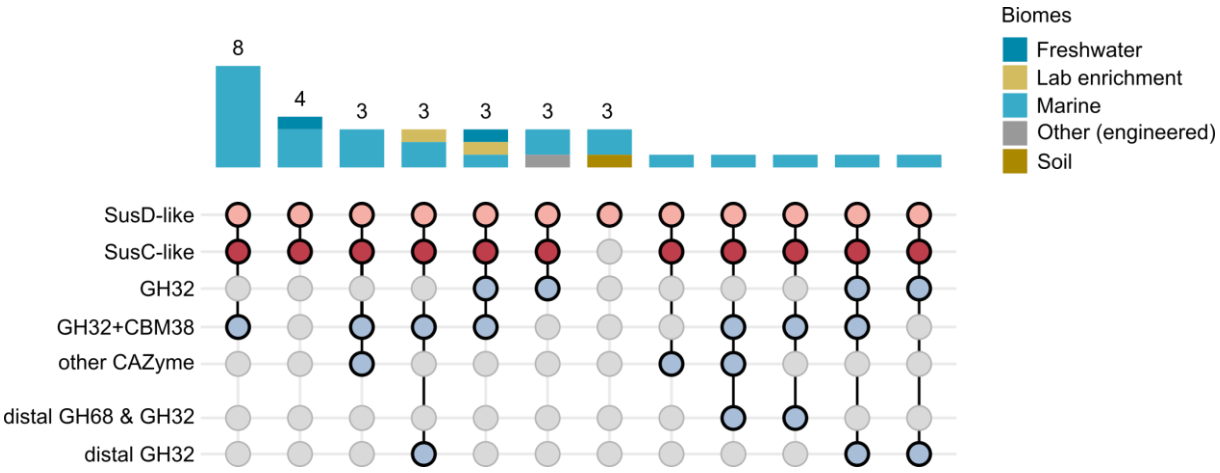

Fig. 4. Gammaproteobacterial SusD-like sequences co-occur with SusC-like TBDTs and GH32s. The GEM and UGCMP databases were searched for gammaproteobacterial SusD-like sequences (4,681 representative genomes). Shown is their genomic context as well as the presence of additional GH68/GH32 pairs or GH32s.

In addition, this analysis revealed that all gammaproteobacterial SusC-like protein sequences with a GH32 context lacked the N-terminal carboxypeptidase D regulatory-like domain (PF13715), as was the case for the *P. distincta* SusC-like protein.

Discussion

According to the current state of research, SusD-like protein/SusC-like TBDT complexes are restricted to the phylum *Bacteroidota*, while *Gammaproteobacteria* (phylum *Pseudomonadota*) import glycans with TBDTs alone. Our study shows an exception to this 'rule'. Here we show that marine *Gammaproteobacteria* such as *P. distincta* use SusC/D-like proteins in fructan utilization.

P. distincta prefers inulin over levan. Consistent with this, *Bacteroidota* from the human intestinal microbiota grew on inulin or levan, but never on both (5). Here, the specificity is mediated by SusC/D-like proteins and the subcellular localization of GH32s (8). The PUL-encoded SusD-like protein ($PdSusD_{PUL}$) of *P. distincta* binds FOS_{INU}, however, with a weak affinity (K_a 43 M^{-1}). Similarly, affinity of the inulin-binding SusD-like protein of *B. thetaiotaomicron* strain Bt-8736 was lower (K_d ~0.5 mM) (8) compared to the levan-binding SusD-like protein of *B. thetaiotaomicron* VPI-5482^T (K_a 1.9-2.3 x 10^4 M^{-1}) (5, 11) and the corresponding SusC/D-like complex (K_d ~1 µM, binding DP8 from levan) (30). Detecting ligand binding can be a challenge. For a non-binding SusD-like protein encoded in xylo-oligosaccharide PUL, it was speculated that binding may require the functional complex with the SusC-like TBDT (37). In *B. thetaiotaomicron* VPI-5482^T, this complex is maintained by interactions between loops of both proteins and moreover, it was demonstrated that binding of FOS from levan spanned both the SusD- and the SusC-like protein (11, 30, 38). These loops and also some residues involved in binding are also conserved in the *P. distincta* proteins using the AF-predicted structures (Fig. S9). On the other hand, it was shown that the deletion of the levan-specific SusD-like protein in *B. thetaiotaomicron* VPI-5482^T still allowed transport via the SusC-like TBDT *in vivo* (30) and that the SusD-like protein alone was binding levan *in vitro* as mentioned before (5, 11). Furthermore, there are so far only a very few reports on marine fructans (21). Low affinity of $PdSusD_{PUL}$ to FOS_{INU} could therefore also be related to the fact that marine fructans differ from terrestrial fructans in their structure. In support of this, $PdSusD_{PUL}$ did not cluster with inulin- or levan-specific sequences from *Bacteroidota* of the human intestine (Fig. S7). However, growth of *P. distincta* and activity of $PdGH32_{PUL}$ on the

inulin and levans used give reason to suspect a high similarity of marine fructan-like glycans to their terrestrial counterpart.

For *Bacteroidota* from the human intestine it was suggested that, compared to levan, small-sized inulin is directly imported by specific SusC/D-like proteins and that these strains lack extracellular GH32s or that extracellular CAZymes were not necessary for growth (8, 39). While our analysis suggests that the PUL-encoded $PdGH32_{PUL}$ of *P. distincta* is presumably tethered to the outer membrane, it could be orientated towards the extracellular space or the periplasm. A periplasmic localization would imply the direct import of inulin, mentioned above (Fig. 5). In this scenario, reduced growth of *P. distincta* on levan could be due to the low abundance of the extracellular distal $PdGH32_{DIS}$, although its activity remains unknown.

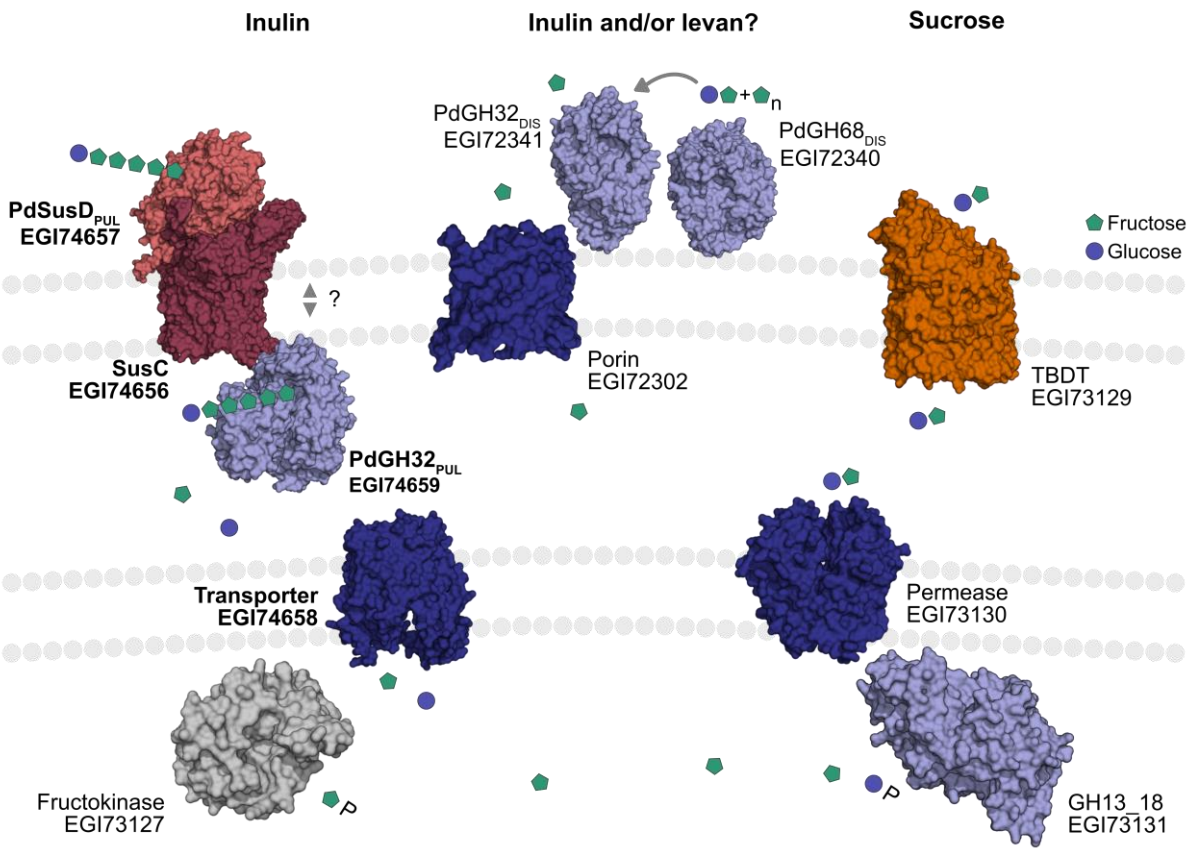

Fig. 5. *Pseudoalteromonas distincta* is adapted to a fructose-based diet. We hypothesize that *P. distincta* imports small-sized inulin directly into the periplasm, where it is degraded by the fructan PUL-encoded $PdGH32_{PUL}$, as previously described for *Bacteroidota* of the human intestine (8). Compared to the fructan PUL-encoded inner membrane sugar transporter EGI74658, the permease EGI74660 of the same PUL was not quantified in any sample. Transporters could be specific for glucose or fructose. In addition, $PdGH32_{PUL}$ could also be orientated towards the extracellular space, releasing fructose, which is more likely to enter via the porin. For further information, please refer to the discussion. The distal

PdGH32$_{DIS}$ presumably recycles fructan-based EPS produced by PdGH68$_{DIS}$. Fructan PUL-encoded proteins are highlighted in bold. EGI73127-31 are encoded in the sucrose PUL (Fig. 1B).

However, the PUL-encoded PdGH32$_{PUL}$ was highly abundant in all subcellular proteomes, e.g., it was the second most abundant protein (7%) in the extracellular proteome of cells grown on plant-derived levan (Supplementary Dataset 2, Fig. 3). Thus, in another scenario, an extracellular orientation of PdGH32$_{PUL}$ would release fructose from fructans, rendering the SusC/D-like proteins unnecessary. Fructose might be imported by porins, one of which was highly abundant in the subcellular proteomes (Fig. 3). Similarly, an exo-cleaving GH32 presumably digests FOS$_{INU}$ extracellularly in *Phocaeicola vulgatus* (formerly *Bacteroides vulgatus*), but the corresponding gene is not co-localized with a *susC/susD* pair (40). The gammaproteobacterial SusC-like proteins identified here lacked the N-terminal carboxypeptidase D regulatory-like domain. The exact function of this domain is unknown, but most PUL-encoded SusC-like TBDTs of *B. thetaiotaomicron* VPI-5482^{T} possess it (31). In the case of the levan-specific SusC-like TBDT (BT_1763) it was shown that the deletion of the carboxypeptidase D regulatory-like domain causes a growth defect (30). It was speculated that it might mediate specificity for the various TonB orthologs.

Nevertheless, *P. distincta* produces the fructan PUL-encoded SusC/D-transport system abundantly, supporting its role in inulin use. The presence of SusC/D-like proteins in *Gammaproteobacteria* underscores that they can use these proteins raising the question of why they are not more frequent and restricted to fructan utilization in this phylum. Overall, it remains speculative, if the only benefit of acquiring a fructan PUL from a putative *Bacteroidota* ancestor was an extended substrate repertoire or if the SusC/D-like tandem came with advantageous aspects. Horizontal transfers of genes or operons related to polysaccharide use have been described before, from marine *Bacteroidota* to *Bacteroidota* of the human intestine (41, 42), but also from a marine flavobacterium (*Bacteroidota*) to *Proteobacteria*, including *Gammaproteobacteria* (43). However, in the latter case, the transfer of *Bacteroidota* alginolytic operons to *Gammaproteobacteria* did not include the SusC/D-like sequences. It is therefore

intriguing, how the fructan-related complex was able to establish in *Gammaproteobacteria*, while others did not.

In general, our analyses demonstrated that fructan PULs are not widely distributed in marine *Gammaproteobacteria*, which confirms the data on the distribution of fructan PULs in bacterial metagenome-assembled genomes from phytoplankton blooms (18). At the same time, GH32s often occur without a PUL context. While *P. distincta* encodes the distal $PdGH32_{DIS}$ and $PdGH68_{DIS}$, an additional GH32-containing PUL may have the advantage that expression is fine-tuned by regulatory elements. This suggests that *P. distincta* is specialized in the utilization of marine fructans, which could include bacterial fructan-based EPS released, e.g., during phage-induced cell lysis, in addition to the still unknown eukaryotic fructan sources. In human intestinal communities, fructan produced by *Lactobacillus reuteri* was used by *B. thetaiotaomicron* (44). Similarly, in marine bacteria that thrive in phytoplankton blooms, it was recently described that *Bacteroidota* may feed on bacterial α-glucan storage glycans (27).

In summary, we discovered SusC/D-like proteins in fructan PULs of marine *Gammaproteobacteria*, which indicates an exchange of protein functions between phyla to support glycan utilization. Our findings reveal that fructose-containing substrates contribute to the marine glycan universe targeted by specialized bacteria such as *P. distincta*.

Materials and Methods

Carbohydrates

Plant-derived levan (Timothy grass) and fructo-oligosaccharides (FOS_{INU}, DP2-8) from inulin (chicory) were purchased from Megazyme/NEOGEN, inulin (chicory) from Alfa Aesar (VWR) and bacterial levan (*Erwinia herbicola*), pullulan and maltotriose from Sigma Aldrich (Merck). Unless otherwise stated, the molecular weights/DP given in the main text correspond to the manufacturer's specifications.

Cultivation of P. distincta

Cells of overnight cultures in marine broth 2216 were pelleted and washed once with an artificial seawater medium (MPM) (45). This suspension was then used to inoculate main cultures of *P. distincta* in MPM supplemented with 0.2% of the respective substrates to an OD_{600nm} of 0.1. Inoculated MPM medium without any carbon source served as a control. Cells were cultivated at 10 °C and 150 rpm (30-260 h).

Proteome analysis

For whole-cell proteomes, *P. distincta* was cultivated in triplicates with 0.2% inulin, plant-derived levan, bacterial levan, FOS_{INU}, pullulan or maltotriose (see section before). In addition, we analyzed the intracellular soluble and extracellular fraction, outer membranes and outer membrane vesicles from inulin- and plant levan-grown cells.

Cells were harvested by centrifugation (15 min, 9,500 x *g*, 4 °C) in the late exponential growth phase. For whole-cell proteomes, cell pellets were resuspended in TE-buffer (10 mM Tris, 10 mM EDTA) and cells were lysed via sonication (4 cycles of 25 s, 30% power, pulse 5, Bandelin Sonopuls HD 2070 ultrasonic homogenizer). Cell debris was removed by centrifugation. Intracellular soluble fractions, outer membranes and outer membrane vesicles were prepared as described before (24, 46, 47). Proteins were extracted from the supernatant using StrataClean beads according to published protocols (48, 49), referred to as the extracellular proteome. Here, 20 µL of bead solution were used to extract 20-25 µg of protein. Otherwise, the Pierce BCA assay determined protein concentration and 25 µg of protein was separated by 1D-SDS-PAGE. Lanes were cut into 10 equal pieces and proteins were digested with trypsin as described before (49). Peptides were subjected to reversed phase chromatography and analyzed using an online-coupled LTQ Orbitrap (whole-cell proteomes) or LTQ XL Orbitrap (subcellular fractions) mass spectrometer equipped with a nanoelectrospray ion source (Thermo Fisher Scientific Inc., Waltham, MA, USA). Data analysis was performed using MaxQuant v. 1.6.0.16 with the integrated Andromeda search engine (50). The false discovery rate (FDR) was set to 1% on peptide and protein level and match between runs was enabled (for subcellular proteomes only between replicates). %riBAQ values (relative intensity-based

absolute quantification giving the relative protein abundance in the sample) were calculated to determine protein abundance per sample. Proteins that were quantified in at least two out of three replicates were considered for calculation of mean values.

For statistical analysis in the case of whole-cell proteomes, the data set was filtered for at least one valid value per protein and missing values were imputed with a small constant. Values were then log-transformed and z-score transformation was used to analyze differential production of proteins. Significant differences were determined by one-way analysis of variance (ANOVA) followed by a Tukey's honestly significant difference (HSD) post hoc test (FDR 5%) in Perseus v. 2.1.0.0 (51, 52) using the filtered, imputed and log-transformed abundance values of proteins. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (53, 54) partner repository, see Data Availability Statement. PSORTb v. 3.0 (55) and DeepLocPro v. 1.0 (56) in combination with SignalP v. 6.0 (57) were used to evaluate suggested subcellular localization of proteins.

DNA cloning

The Gibson assembly strategy (58) was used to clone genes into the pET28a vector with primers listed in Table S1. Primers were designed to produce *P. distincta* proteins as recombinant N-terminal His_6-tagged constructs, removing predicted signal peptides from sequences (57). Genes and the pET28a vector were amplified using the PrimeSTAR GXL DNA Polymerase (Takara). The NEBuilder® HiFi DNA Assembly Master Mix (NEB) was used for assembly following the manufacturer's protocol. Sequencing (Eurofins) confirmed sequence identity. For protein production, plasmids were transformed into competent *Escherichia coli* BL21(DE3) cells.

Protein production & purification

E. coli BL21(DE3) clones, harbouring the corresponding plasmids, were grown in 1 L lysogeny broth (LB) containing 30 µg mL^{-1} kanamycin at 37 °C and 120 rpm. Medium was cooled down before isopropyl β-D-1-thiogalactopyranoside (IPTG) induction (0.3 or 1 mM) and cells were grown overnight at 20 °C at 120 rpm. In the case of CAZyme activity analysis, proteins were

produced using the ZYP5052 autoinduction medium (59). Here, cells were cultivated at 37°C until an OD_{600nm} 0.6-0.8 and then at 20 °C overnight. Cells were harvested by centrifugation and pellets were stored at -20 °C until further treatment. A chemical lysis protocol (60) was used to disrupt the cells (resuspension buffer: 50 mM Tris-HCl pH 8.0, 25% sucrose, ~10 mg lysozyme; lysis buffer: 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1% sodium DCA, 1% Triton-X). For CAZyme activity tests, cells were lysed via ultrasonication on ice (Bandelin Sonoplus HD2200, KE76 sonotrode), 2 x 3 min, 1 min pause, 50% power, 0.5 pulse.

For protein purification, 5 mL HisTrap HP columns were used, equilibrated in buffer A (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole) and charged with 0.1 M $NiSO_4$. The proteins were eluted using a linear gradient of 0-100% buffer B (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 500 mM imidazole) within 60 min. The purity of the sample was increased by size-exclusion chromatography (SEC) using a Superdex 200 16/60 column (Cytiva) in buffer C (CAZyme activity: 25 mM Tris-HCl pH 8.0, 50 mM NaCl, otherwise: 20 mM Tris-HCl pH 8.0, 200-250 mM NaCl). Protein concentration was determined on a Nanodrop Spectrophotometer using the molecular weight and extinction coefficient (61). Prior to SEC, proteins were concentrated at 3,500 rpm and 4 °C using Amicon Ultra centrifugal filter units (Merck) or using a stirred cell ultrafiltration device, both 10 kDa cut-off.

Reducing sugar assay (RSA)

The degradation of different oligo- and polysaccharides by CAZymes was carried out using final concentrations of 2 mg mL^{-1} substrate and 30 µg mL^{-1} of CAZyme in 1 x PBS, pH 6.0. The enzymatic reactions were incubated for 24 h at room temperature. For the quantification of reducing ends released upon degradation, 20 µL of the reactions were mixed with 20 µL DNS-reagent (0.4 M NaOH, 30% potassium sodium tartrate, 1% 3,5-dinitrosalicylic acid; DNS) and incubated at 100 °C for 15 min. dH_20 (180 µL) was added and 200 µL of the mixtures were transferred to a microtiter plate. The absorption at 540 nm was measured using a Tecan infinite 200 pro.

Fluorophore-assisted carbohydrate electrophoresis (FACE)

The enzymatic reactions (10 µL) were lyophilized, resuspended in ANTS (0.1 M in DMSO, 4 µL, Invitrogen™) and $NaCNBH_3$ (1 M in DMSO, 4 µL) and incubated at 37 °C overnight. The labeled samples were mixed with loading dye (62 mM Tris pH 6.8, 0.014% bromophenol blue, 10% glycerol, 8 µL) and separated at 400 V on a 30% polyacrylamide running gel with a 4% stacking gel for about 1h under consistent cooling. Imaging was performed using a 312 nm UV table.

Isothermal titration calorimetry (ITC)

$PdSusD_{PUL}$ was dialysed against 50 mM Tris pH 8.0, 100 mM NaCl at 4 °C for at least 48 h. ITC analysis was performed using a MicroCal™ ITC 200 machine. The dialyses buffer was used to dissolve the substrates and to wash the ITC sample cell. All samples were centrifuged before analysis. 63.5 mM of FOS_{INU} (calculated as DP5) were injected into 160.97 µM of $PdSusD_{PUL}$. The experiment was performed in triplicate. Injections with 20 mM of inulin (calculated as DP17), 5 mM of plant-derived levan, 7.5 µM of bacterial levan and 20 mM of sucrose or fructose detected no signal. Since little heat was released during titrations when the concentration of FOS_{INU} was increased to 31.7 mM, we tested even higher concentrations of FOS_{INU}. The following settings were selected: cell temperature 20 °C (293.15 K), reference power 10 µCal/s, stirring speed 500 rpm, filter period 1 sec, injection spacing 200 sec, 2 µL injection volume (first injection 0.3 µL), 18 injections in total. As a control, FOS_{INU} was titrated into buffer and buffer into buffer. Due to the fact that the low affinity resulted in a C-value of 0.01 (62) and to the best of our knowledge SusD-like proteins bind in a 1:1 ratio, the n-value was fixed to 1. MICROCAL ORIGIN v.7 was used to analyze the data applying a single-site binding model.

Protein structure prediction and analysis

AlphaFold2-predicted structures were downloaded from the AlphaFold Protein Structure DB (63), which is integrated in the Uniprot knowledgebase (64) and visualized using PyMOL Molecular Graphics System v2.5.4 (Schrödinger, LLC.).

Phylogenetic analysis

A blast of $PdSusD_{PUL}$ against the UniProtKB and Reference Proteomes retrieved 1,750 unique sequences. Additional SusD-like protein sequences encoded in fructan PULs from *Bacteroidota* of the human intestine (with suspected or known specificity for inulin or levan (5, 8, 11)) as well as SusD-like protein-coding sequences from PULs of *Zobellia galactanivorans* $Dsij^T$ with known substrate specificity (other than fructans) were added to the dataset. Thus, in total, 1776 sequences were aligned using MAFFT v. 7 (65) using the L-INS-i algorithm. The obtained multiple sequence alignment was visualized and processed in Jalview v. 11.0 (66). Short sequences, signal peptides and non-aligned regions were removed. Sequences were filtered at 90 % identity and finally, 1055 sequences with 410 positions were used for the phylogeny. The phylogenetic tree was built by the maximum likelihood method (67) using RAxML v. 8.2.4 (68). The best evolutive model was selected submitting the alignment to IQ-TREE web server (http://iqtree.cibiv.univie.ac.at; (69). The LG matrix as evolutive model (70) and a discrete Gamma distribution to model evolutionary rate differences among sites (4 categories) were used for the calculation. The rate variation model allowed for some sites to be evolutionarily invariable. The reliability of the trees was tested by bootstrap analysis using 1000 resamplings of the dataset (71). The tree was visualized using iTOL v. 5 (72) and midpoint rooted.

Co-occurrence analysis

Gammaproteobacterial genomes of the unified genome catalog of marine prokaryotes (UGCMP) (73) and the Genomes from Earth's Microbiomes (GEM) catalog (74) were dereplicated using dRep (v3.4.2, flags: -l 0 -comp 50 -con 10 -pa 0.95 --checkM_method taxonomy_wf) (75). Resulting 4,681 genomes were reannotated using prokka (v1.14.6) (76) and screened for SusD-like homologs using hmmsearch (v3.3.2, gathering threshold) (77) with PFAM model PF07980.15 (SusD_RagB), PF12741.11 (SusD-like), PF12771.11 (SusD-like_2) and PF14322.10 (SusD-like_3) (78).

In order to examine polysaccharide utilization-related functions around *susD-like* genes, relevant functions were predicted using a five-gene sliding window. For TBDT annotation, sequences were compared to PF13715.11 (CarboxypepD_reg-like domain), PF07715.18 (Plug) and PF00593.27 (TonB_dep_Rec) using hmmsearch (gathering threshold). Sequences containing PF07715.18 and PF00593.27 were kept for further analysis. TBDT also fitting the gathering threshold for TIGRFAM profile TIGR04056 (OMP_RagA_SusC) (79) were classified as SusC-like proteins. CAZyme families were assigned with hmmscan vs. dbCAN-HMMdb-V12 and dbCAN-sub as well as diamond blastp (v2.1.1.155, e-value cut-off 1E-102) against CAZyDB.07262023, all provided by dbCAN (80). Results were then filtered using the hmmscan-parser.sh script with an e-value cutoff of 1E-15 and a minimum coverage of 0.35. CAZyme families predicted by at least two tools were kept for further analysis, not considering GT and AA assignments. Results were visualized with the ComplexUpset R package (v1.3.5) (81).

Furthermore, the 4,681 representative gammaproteobacterial genomes were searched for GH32 family encoding genes with hmmsearch (v3.3.2, option settings: -E 1 --domE 1 --incE 0.01 --incdomE 0.03) using GH32 HMM profiles from dbCAN-HMMdb-V12 and dbCAN-sub. Hits were further assigned to a final CAZyme family as described above. Their genetic neighborhood 5 genes up- and downstream was reconstructed similar to the predictions mentioned above, including hmmscan (gathering threshold) against the full Pfam-A database (as of 20240508).

Data Availability

The proteomic data were deposited on the PRIDE archive accessible with the identifiers PXD054951 (whole-cell proteomes) and PXD054992 (subcellular proteomes). Data will be released upon acceptance of the manuscript in a peer-reviewed journal.

Acknowledgements

This project was supported by the German Research Foundation (DFG) within the research unit FOR 2406 ‘Proteogenomics of Marine Polysaccharide Utilization’ (POMPU) by grants to

Dörte Becher (BE 3869/4-3), Uwe Bornscheuer (BO 1862/17-3), Jan-Hendrik Hehemann (HE 7217/2-3), and Thomas Schweder (SCHW 595/10-3, SCHW 595/11-3). The authors thank Jana Matulla, Theresa Dutschei, Julia-Maria Seeßelberg and Marcella Consolo for support in the lab as well as Sebastian Grund for LC-MS/MS measurements. We are grateful for having access to the CristalO platform (FR2424, Station Biologique de Roscoff), which is part of the core facility networks Biogenouest and EMBRC-France, and to the P11 beamline at DESY (Hamburg, Germany) and thank the staff for their support on site. MKZ thanks the DFG (SCHW 595/10-3), which enabled a research stay in Roscoff.

Author contributions

MKZ and TS designed and directed the study. MKZ, AB, VS, MT, FU, EFB, LC, AJ, AS, FH, DJ, MJ and MC performed research. MKZ, DB, EFB, TB, MH and MC analyzed the data. MD, UB, DB, JHH, MC and TS provided resources. MKZ wrote the paper, which was edited and approved by all authors.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Versluys M, Kirtel O, Toksoy Öner E, Van den Ende W. The fructan syndrome: Evolutionary aspects and common themes among plants and microbes. Plant Cell Environ. 2018;41(1):16-38.

2. Kumar AS, Mody K, Jha B. Bacterial exopolysaccharides - a perception. J Basic Microbiol. 2007;47(2):103-17.

3. Öner ET, Hernández L, Combie J. Review of Levan polysaccharide: From a century of past experiences to future prospects. Biotechnol Adv. 2016;34(5):827-44.

4. Zhang WL, Xu W, Ni DW, Dai QY, Guang CE, Zhang T, et al. An overview of levan-degrading enzyme from microbes. Appl Microbiol Biotechnol. 2019;103(19):7891-902.

5. Sonnenburg ED, Zheng HJ, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, et al. Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell. 2010;141(7):1241-U256.

6. Ernits K, Eek P, Lukk T, Visnapuu T, Alarnäe T. First crystal structure of an endo-levanase - the BT1760 from a human gut commensal *Bacteroides thetaiotaomicron*. Sci Rep. 2019;9.

7. Mardo K, Visnapuu T, Vija H, Aasamets A, Viigand K, Alamae T. A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian Gut Commensal *Bacteroides thetaiotaomicron* Cleaves Not Only Various Bacterial Levans, but Also Levan of Timothy Grass. Plos One. 2017;12(1):e0169989.

8. Joglekar P, Sonnenburg ED, Higginbottom SK, Earle KA, Morland C, Shapiro-Ward S, et al. Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in *Bacteroides thetaiotaomicron* Strains. mSphere. 2018;3(3).

9. Anderson KL, Salyers AA. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J Bacteriol. 1989;171(6):3192-8.

10. Bjursell MK, Martens EC, Gordon JI. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. J Biol Chem. 2006;281(47):36269-79.

11. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature. 2017;541(7637):407-11.

12. Bolam DN, van den Berg B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. Curr Opin Struct Biol. 2018;51:35-43.

13. Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B. *Bacteroidetes* use thousands of enzyme combinations to break down glycans. Nat Commun. 2019;10(1):2043.

14. Li AX, Guo LZ, Lu WD. Alkaline inulinase production by a newly isolated bacterium Marinimicrobium sp LS-A18 and inulin hydrolysis by the enzyme. World J Microbiol Biotechnol. 2012;28(1):81-9.

15. Gao L, Chi Z, Sheng J, Wang L, Li J, Gong F. Inulinase-producing marine yeasts: evaluation of their diversity and inulin hydrolysis by their crude enzymes. Microb Ecol. 2007;54(4):722-9.

16. Abd El Aty AA, Wehaidy HR, Mostafa FA. Optimization of inulinase production from low cost substrates using Plackett-Burman and Taguchi methods. Carbohydr Polym. 2014;102:261-8.

17. Liebl W, Brem D, Gotschlich A. Analysis of the gene for beta-fructosidase (invertase, inulinase) of the hyperthermophilic bacterium *Thermotoga maritima*, and characterisation of the enzyme expressed in *Escherichia coli*. Appl Microbiol Biotechnol. 1998;50(1):55-64.

18. Wang FQ, Bartosik D, Sidhu C, Siebers R, Lu DC, Trautwein-Schult A, et al. Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides. Microbiome. 2024;12(32).

19. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH, et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. Microbiome. 2023;11(1):77.

20. Alberto F, Jordi E, Henrissat B, Czjzek M. Crystal structure of inactivated invertase in complex with the trisaccharide substrate raffinose. Biochem J. 2006;395:457-62.

21. Okutani K. Structural Investigation of the Fructan from Marine Bacterium Nam-1. Bull Jap Soc Sci Fish. 1982;48(11):1621-5.

22. Daly G, Decorosi F, Viti C, Adessi A. Shaping the phycosphere: Analysis of the EPS in diatom-bacterial co-cultures. J Phycol. 2023;59(4):791-7.

23. Gobet A, Barbeyron T, Matard-Mann M, Magdelenat G, Vallenet D, Duchaud E, et al. Evolutionary Evidence of Algal Polysaccharide Degradation Acquisition by *Pseudoalteromonas carrageenovora* 9^T to Adapt to Macroalgal Niches. Front Microbiol. 2018;9.

24. Dürwald A, Zühlke MK, Schlüter R, Gebbe R, Bartosik D, Unfried F, et al. Reaching out in anticipation: bacterial membrane extensions represent a permanent investment in polysaccharide sensing and utilization. Environ Microbiol. 2021;23(6):3149-63.

25. Keith J, Wiley B, Ball D, Arcidiacono S, Zorfass D, Mayer J, et al. Continuous culture system for production of biopolymer levan using *Erwinia herbicola*. Biotechnol Bioeng. 1991;38(5):557-60.

26. Vijn I, Smeekens S. Fructan: More than a reserve carbohydrate? Plant Physiol. 1999;120(2):351-9.

27. Beidler I, Steinke N, Schulze T, Sidhu C, Bartosik D, Zühlke MK, et al. Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle. Nat Commun. 2024;15(1).

28. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res. 2022;50(W1):W210-W5.

29. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(4):486-501.

30. Gray DA, White JBR, Oluwole AO, Rath P, Glenwright AJ, Mazur A, et al. Insights into SusCD-mediated glycan import by a prominent gut symbiont. Nat Commun. 2021;12(1):44.

31. Pollet RM, Martin LM, Koropatkin NM. TonB-dependent transporters in the *Bacteroidetes*: Unique domain structures and potential functions. Mol Microbiol. 2021;115(3):490-501.

32. Pouyez J, Mayard A, Vandamme AM, Roussel G, Perpète EA, Wouters J, et al. First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: discovery of an extra-pocket in the catalytic domain responsible for its endo-activity. Biochimie. 2012;94(11):2423-30.

33. Nagem RA, Rojas AL, Golubev AM, Korneeva OS, Eneyskaya EV, Kulminskaya AA, et al. Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. J Mol Biol. 2004;344(2):471-80.

34. Lee JH, Kim KN, Choi YJ. Identification and characterization of a novel inulin binding module (IBM) from the CFTase of *Bacillus macerans* CFC1. FEMS Microbiol Lett. 2004;234(1):105-10.

35. Park J, Kim MI, Park YD, Shin I, Cha J, Kim CH, et al. Structural and Functional Basis for Substrate Specificity and Catalysis of Levan Fructotransferase. J Biol Chem. 2012;287(37):31233-41.

36. Yanase H, Iwata M, Nakahigashi R, Kita K, Kato N, Tonomura K. Purification, Crystallization, and Properties of the Extracellular Levansucrase from *Zymomona mobilis*. Biosci Biotech Bioch. 1992;56(8):1335-7.

37. Tauzin AS, Wang Z, Cioci G, Li XQ, Labourel A, Machado B, et al. Structural and Biochemical Characterization of a Nonbinding SusD-Like Protein Involved in Xylooligosaccharide Utilization by an Uncultured Human Gut *Bacteroides* Strain. Msphere. 2022;7(5).

38. White JBR, Silale A, Feasey M, Heunis T, Zhu YL, Zheng H, et al. Outer membrane utilisomes mediate glycan uptake in gut *Bacteroidetes*. Nature. 2023.

39. Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. Nature. 2016;533(7602):255-9.

40. Bolam DN, Sonnenburg JL. Mechanistic insight into polysaccharide use within the intestinal microbiota. Gut Microbes. 2011;2(2):86-90.

41. Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010;464(7290):908-12.

42. Pudlo NA, Pereira GV, Parnami J, Cid M, Markert S, Tingley JP, et al. Diverse events have transferred genes for edible seaweed digestion from marine to human gut bacteria. Cell Host Microbe. 2022;30(3):314-+.

43. Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine

Flavobacteriia to their independent transfers to marine Proteobacteria and human gut *Bacteroides*. Environ Microbiol. 2012;14(9):2379-94.

44. van Bueren AL, Saraf A, Martens EC, Dijkhuizen L. Differential Metabolism of Exopolysaccharides from Probiotic Lactobacilli by the Human Gut Symbiont Bacteroides thetaiotaomicron. Appl Environ Microbiol. 2015;81(12):3973-83.

45. Schut F, de Vries EJ, Gottschal JC, Robertson BR, Harder W, Prins RA, et al. Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. Appl Environ Microbiol. 1993;59(7):2150-60.

46. Frias A, Manresa A, de Oliveira E, López-Iglesias C, Mercade E. Membrane vesicles: a common feature in the extracellular matter of cold-adapted Antarctic bacteria. Microb Ecol. 2010;59(3):476-86.

47. McBroom AJ, Johnson AP, Vemulapalli S, Kuehn MJ. Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability. J Bacteriol. 2006;188(15):5385-92.

48. Bonn F, Otto A. Protein Enrichment from Highly Dilute Samples with StrataClean. Methods Mol Biol. 2018;1841:11-8.

49. Bonn F, Bartel J, Büttner K, Hecker M, Otto A, Becher D. Picking vanished proteins from the void: how to collect and ship/share extremely dilute proteins in a reproducible and highly efficient manner. Anal Chem. 2014;86(15):7421-7.

50. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72.

51. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods. 2016;13(9):731.

52. Tyanova S, Cox J. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. Methods Mol Biol. 2018;1711:133-48.

53. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 2016;44(D1):D447-56.

54. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543-D52.

55. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-15.

56. Moreno J, Nielsen H, Winther O, Teufel F. Predicting the subcellular location of prokaryotic proteins with DeepLocPro. bioRxiv Preprint January 04, 2024. 2024;Available from: https://doi.org/10.1101/2024.01.04.574157.

57. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023-5.

58. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343-5.

59. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif. 2005;41(1):207-34.

60. Hehemann JH, Smyth L, Yadav A, Vocadlo DJ, Boraston AB. Analysis of Keystone Enzyme in Agar Hydrolysis Provides Insight into the Degradation of a Polysaccharide from Red Seaweeds. J Biol Chem. 2012;287(17):13985-95.

61. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531-52.

62. Wiseman T, Williston S, Brandts JF, Lin LN. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. Anal Biochem. 1989;179(1):131-7.

63. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439-D44.

64. Consortium TU. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480-D9.

65. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-66.

66. Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. Bioinformatics. 2004;20(3):426-7.

67. Felsenstein J. Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach. J Mol Evol. 1981;17(6):368-76.

68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

69. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44(W1):W232-W5.

70. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307-20.

71. Felsenstein J. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution. 1985;39(4):783-91.

72. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.

73. Nishimura Y, Yoshizawa S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. Sci Data. 2022;9(1):305.

74. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39(4):499-509.

75. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11(12):2864-8.

76. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

77. Eddy SR. Accelerated Profile HMM Searches. PLoS Comp Biol. 2011;7(10):e1002195.

78. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D9.

79. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31(1):371-3.

80. Zheng J, Hu B, Zhang X, Ge Q, Yan Y, Akresi J, et al. dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. Nucleic Acids Res. 2023;51(D1):D557-D63.

81. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. Ieee T Vis Comput Gr. 2014;20(12):1983-92.