

Gene Expression Analysis of *Zobellia galactanivorans* during the Degradation of Algal Polysaccharides Reveals both Substrate-Specific and Shared Transcriptome-Wide Responses

François Thomas, Philippe Bordron, Damien Eveillard, Gurban Michel

► **To cite this version:**

François Thomas, Philippe Bordron, Damien Eveillard, Gurban Michel. Gene Expression Analysis of *Zobellia galactanivorans* during the Degradation of Algal Polysaccharides Reveals both Substrate-Specific and Shared Transcriptome-Wide Responses. *Frontiers in Microbiology*, Frontiers Media, 2017, 8, pp.article 1808. <10.3389/fmicb.2017.01808>. <hal-01613034>

HAL Id: hal-01613034

<https://hal.archives-ouvertes.fr/hal-01613034>

Submitted on 16 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Gene Expression Analysis of *Zobellia galactanivorans* during the Degradation of Algal Polysaccharides Reveals both Substrate-Specific and Shared Transcriptome-Wide Responses

François Thomas^{1*}, Philippe Bordron^{2,3,4}, Damien Eveillard⁵ and Gurvan Michel^{1*}

¹ Sorbonne Universités, UPMC Univ Paris 06, Centre National de la Recherche Scientifique, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, Roscoff, France, ² Sorbonne Universités, UPMC Univ Paris 06, Centre National de la Recherche Scientifique, FR2424, Analysis and Bioinformatics for Marine Science, Station Biologique de Roscoff, Roscoff, France, ³ Mathomics, Center for Mathematical Modeling, Universidad de Chile, Santiago, Chile, ⁴ Center for Genome Regulation (Fondap 15090007), Universidad de Chile, Santiago, Chile, ⁵ Université de Nantes, Laboratoire des Sciences du Numérique de Nantes, Centre National de la Recherche Scientifique, ECN, IMTA, Nantes, France

OPEN ACCESS

Edited by:

Catherine Ayn Brissette,
University of North Dakota,
United States

Reviewed by:

Thomas Schweder,
University of Greifswald, Germany
Matthias Wietz,
University of Oldenburg, Germany

*Correspondence:

François Thomas
francois.thomas@sb-roscoff.fr
Gurvan Michel
gurvan.michel@sb-roscoff.fr

Specialty section:

This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

Received: 04 July 2017

Accepted: 05 September 2017

Published: 21 September 2017

Citation:

Thomas F, Bordron P, Eveillard D and Michel G (2017) Gene Expression Analysis of *Zobellia galactanivorans* during the Degradation of Algal Polysaccharides Reveals both Substrate-Specific and Shared Transcriptome-Wide Responses. *Front. Microbiol.* 8:1808. doi: 10.3389/fmicb.2017.01808

Flavobacteriia are recognized as key players in the marine carbon cycle, due to their ability to efficiently degrade algal polysaccharides both in the open ocean and in coastal regions. The chemical complexity of algal polysaccharides, their differences between algal groups and variations through time and space, imply that marine flavobacteria have evolved dedicated degradation mechanisms and regulation of their metabolism during interactions with algae. In the present study, we report the first transcriptome-wide gene expression analysis for an alga-associated flavobacterium during polysaccharide degradation. *Zobellia galactanivorans* Dsij^T, originally isolated from a red alga, was grown in minimal medium with either glucose (used as a reference monosaccharide) or one selected algal polysaccharide from brown (alginate, laminarin) or red algae (agar, porphyran, ι- or κ-carrageenan) as sole carbon source. Expression profiles were determined using whole-genome microarrays. Integration of genomic knowledge with the automatic building of a co-expression network allowed the experimental validation of operon-like transcription units. Differential expression analysis revealed large transcriptomic shifts depending on the carbon source. Unexpectedly, transcriptomes shared common signatures when growing on chemically divergent polysaccharides from the same algal phylum. Together with the induction of numerous transcription factors, this hints at complex regulation events that fine-tune the cell behavior during interactions with algal biomass in the marine environment. The results further highlight genes and loci that may participate in polysaccharide utilization, notably encoding Carbohydrate Active enZymes (CAZymes) and glycan binding proteins together with a number of proteins of unknown function. This constitutes a set of candidate genes potentially representing

new substrate specificities. By providing an unprecedented view of global transcriptomic responses during polysaccharide utilization in an alga-associated model flavobacterium, this study expands the current knowledge on the functional role of flavobacteria in the marine carbon cycle and on their interactions with algae.

Keywords: polysaccharide utilization locus, flavobacteria-algae interactions, regulation of gene expression, operon, microarray

INTRODUCTION

Polysaccharides account for a large fraction of the standing stock of organic matter in marine environments. These polysaccharides arise mainly from primary production by phytoplankton in the open ocean (Field et al., 1998), with an added significant contribution of macroalgae in coastal ecosystems (Gattuso et al., 1998). In both micro- and macroalgae, polysaccharides can constitute more than 50% of the dry weight in the form of carbon storage compounds or cell wall constituents, and can be exuded as extracellular material (Kloareg and Quatrano, 1988; Mykkestad, 1995; Biddanda and Benner, 1997). Microbial degradation of these diverse and abundant polysaccharide sources is regarded as a crucial bottleneck in the marine carbon cycle, allowing transfer of organic matter to higher trophic levels (Azam and Malfatti, 2007; Gasol et al., 2008; Buchan et al., 2014). Marine algal polysaccharides differ largely from those found in terrestrial plants, with respect to the nature of their monosaccharide units and the presence of sulfated motifs (Popper et al., 2011; De Jesus Raposo et al., 2013; Ficko-Blean et al., 2014). This implies that marine bacteria have evolved dedicated degradation mechanisms to utilize the polysaccharides found in their respective environments. However, knowledge on the bacterial degradation of algal polysaccharides and the regulation of these processes is still scarce in comparison to terrestrial polysaccharides.

Recent studies have shown that marine polysaccharides can be used by diverse specialized taxa, including members of *Bacteroidetes*, *Gammaproteobacteria*, *Planctomycetes*, and *Verrucomicrobia* (Martinez-Garcia et al., 2012; Teeling et al., 2012; Lage and Bondoso, 2014; Wietz et al., 2015). In particular, marine *Bacteroidetes* are recognized as key players in the degradation of high molecular weight compounds (Kirchman, 2002; Thomas et al., 2011b). Representatives of the *Flavobacteriia* class are often associated to phytoplankton blooms and macroalgal surfaces and are well known for their ability to utilize a variety of algal polysaccharides (Teeling et al., 2012; Williams et al., 2013; Martin et al., 2014). A decade ago, genome sequencing of the first marine *Flavobacteriia*, *Gramella forsetii* KT0803^T isolated from surface seawater, revealed a substantial suite of carbohydrate active enzymes (CAZymes) such as, 42 glycoside hydrolases (GH), six polysaccharide lyases (PL), eight carbohydrate esterases (CE), and two sulfatases predicted to act on polysaccharides (Bauer et al., 2006). Since then, analysis of several additional genomes from algae-associated flavobacteria uncovered an even greater abundance of CAZyme and sulfatase genes. This includes genomes of

bacterial isolates from phytoplankton such as, *Cellulophaga algicola* IC166^T (total predicted GH + PL + CE + Sulfatases = 101 genes; Abt et al., 2011) and *Polaribacter* sp. Hel1_85 (104 genes; Xing et al., 2015), as well as isolates from macroalgae such as, *Formosa agariphila* M-2Alg 35-1^T (146 genes; Mann et al., 2013) and *Zobellia galactanivorans* Dsij^T (236 genes; Barbeyron et al., 2016b). Furthermore, CAZyme-encoding genes are frequently organized in Polysaccharide Utilization Loci (PUL), which are gene clusters typical to *Bacteroidetes* that encode enzymes, binding proteins and transporters required for the breakdown and uptake of complex carbohydrates (Grondin et al., 2017). This genomic knowledge points at a specialization of marine *Flavobacteriia* toward polysaccharide utilization in the environment. However, transcriptomic profiling is required to understand under which conditions these degradation capacities are expressed and the regulations involved, and ultimately deepen our comprehension of algae-flavobacteria interactions. To date, transcriptome-wide gene expression studies on marine *Flavobacteriia* have mainly focused on light-induced responses associated with proteorhodopsin-enhanced growth in members of the genus *Dokdonia* (Kimura et al., 2011; Gómez-Consarnau et al., 2016) and on the response of *Cellulophaga baltica* to viral infection (Howard-Varona et al., 2017). A first transcriptome-wide profiling was reported for the surface seawater isolate *Gramella flava* JLT2011^T, to characterize its response to the polysaccharides xylan and pectin (Tang et al., 2017). To our knowledge, transcriptome-wide responses linked to polysaccharide degradation have never been investigated in flavobacteria isolates from algae and are still understudied even in other alga-associated bacterial phyla. For instance, Zhu et al. (2016) recently reported a complete gene expression profiling of *Bacillus weihaiensis* Alg07, a *Firmicutes* isolated from rotting seaweed. Using RNA-seq, they showed that *B. weihaiensis* overexpresses several genes involved in alginate metabolism, including genes encoding degradation enzymes, putative transporters, and regulators (Zhu et al., 2016).

In the present study, we explored transcriptome-wide gene expression of *Z. galactanivorans* Dsij^T grown on a diverse set of algal polysaccharides. Members of the *Zobellia* genus are marine flavobacteria often found on the surface of healthy green, red, and brown macroalgae (Hollants et al., 2013; Martin et al., 2015; Marzinelli et al., 2015). Among the five currently described *Zobellia* species, *Z. galactanivorans* Dsij^T has become a model organism for polysaccharide degradation by marine flavobacteria. Originally isolated from the red alga *Delesseria sanguinea* (Hudson) J. V. Lamouroux 1813 (Barbeyron et al., 2001), it can degrade a large variety of

algal polysaccharides. Sequencing of its genome revealed a number of adaptive traits for interactions with algae, such as, consumption of seaweed exudates, potential resistance to algal defense and a huge arsenal of CAZymes and sulfatases to degrade polysaccharides (Barbeyron et al., 2016b). Detailed biochemical studies have begun to unveil complex enzymatic systems for the degradation of agars (Hehemann et al., 2012) and carrageenans (Barbeyron et al., 1998; Rebuffet et al., 2010) from red algae and alginate (Thomas et al., 2012, 2013; Zhu et al., 2017) and laminarin (Labourel et al., 2014, 2015) from brown algae.

Here, we report the results of complete gene expression profiling for *Z. galactanivorans* Dsij^T during the utilization of algal polysaccharides. This approach allowed the prediction of operon-like genomic units throughout the chromosome. It further revealed polysaccharide-responsive regulons and new candidate genes for substrate recognition, transport, and degradation.

METHODS

Substrates, Bacterial Strain, and Culture Conditions

The substrates tested for growth were glucose (Merck), alginate (Sigma), laminarin (Degussa), ι -carrageenan (Sanofi BioIndustries), κ -carrageenan (extracted from *Euचेuma cottonii*, gift from Goëmar), agar (Amresco), or porphyran (extracted from *Porphyra umbilicalis* as in Correc et al., 2011). ¹H NMR analysis of the self-extracted porphyran and κ -carrageenan confirmed the purity and the motif composition of the polysaccharides (Correc et al., 2011; Préchoux et al., 2016). The type strain *Z. galactanivorans* Dsij^T (Barbeyron et al., 2001) was routinely grown from glycerol stocks in Zobell medium 2216E (Zobell, 1941) at 20°C. Cells were pelleted (5 min, 4,000 g) and resuspended in saline solution to remove traces of organic substrates. This suspension was inoculated (1/50 ratio) in triplicate flasks containing marine mineral medium (Thomas et al., 2011a) amended with 2 g.L⁻¹ of one substrate as sole carbon source. Cultures were incubated at 20°C, 180 rpm. After 48 h, aliquots (5 ml) were centrifuged (10 min, 4,000 g) and cell pellets were resuspended in 200 μ l Trizol reagent (Sigma), frozen in liquid nitrogen and stored at -80°C until RNA extraction.

RNA Preparation and cDNA Synthesis

Samples were directly transferred from -80 to a 65°C water bath for 15 min to lyse cells and mixed with 50 μ l chloroform for 5 min at room temperature. The aqueous phase was recovered after centrifugation for 15 min at 13,000 g, 4°C. Total RNA was purified using RNeasy Mini Kit (Qiagen) with on-column DNase I treatment following the manufacturer's instructions, and eluted in 15 μ l RNase-free water. RNA concentration was quantified on a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Inc.). Concentrations and ratios A_{260/280} and A_{260/230} are given in Supplementary Table 1. RNA integrity was confirmed by 0.8% agarose gel electrophoresis. The absence of genomic DNA was confirmed by PCR on

purified RNA. The cDNAs were synthesized from 10 μ g total RNA using the Superscript Double Stranded cDNA kit (Invitrogen), purified by phenol:chloroform:isoamyl alcohol extraction (v/v, 25:24:1), recovered by ethanol precipitation and resuspended in 15 μ l nuclease-free water. The cDNA quality was confirmed on a Bioanalyzer (Agilent) using the RNA 6000 Nano kit (Supplementary Figure 1).

Microarray Analysis

The transcriptomic analysis was performed by PartnerChip (Evry, France) on a custom microarray (4*72k, Roche Nimblegen). The whole genome of *Z. galactanivorans* Dsij^T (4,482 CDS + 2,595 intergenic regions) was covered by 72,000 oligonucleotides with a 60-bp average size. In total, the pangenomic microarray comprised 52,389 and 19,345 probes targeting genic and intergenic regions, respectively. It was impossible to design probes for 256 coding regions and 1,556 non-coding regions, either because sequences were too similar to ensure specificity or the target regions were too short. Overall, the microarray covered 95% of the genes of *Z. galactanivorans*, with 11.6 probes per gene on average. cDNA from each sample (1 μ g) was labeled with Cyanine 3, hybridized on the microarray in a Nimblegen station and washed three times. Cyanine 3 fluorescence was scanned on a MS200 scanner (Roche Nimblegen). Raw intensities were obtained from the NimbleScan software (Nimblegen). The ANAIS web interface (Simon and Biot, 2010) was used to normalize data and detect significantly expressed genes, i.e., genes expressed more than random probes included on the chip ($p < 0.01$). Raw and normalized data were deposited on the NCBI GEO repository under accession GSE99940. Biological triplicate data obtained with each polysaccharide were compared to glucose condition using empirical Bayes statistics implemented in the *eBayes* function from the *limma* package v 3.26.9 in R 3.2.2 (R Core Team, 2013) and Bonferroni correction for multiple tests. Genes with an adjusted p -value FWER ≤ 0.05 were considered as differentially expressed. Non-metric multidimensional scaling (NMDS) was performed on a Morisita-Horn dissimilarity matrix obtained from normalized expression data with the *vegan* package (Oksanen et al., 2013). Sets of differentially expressed genes were compared using UpSetR (Lex et al., 2014).

Detection of Transcriptional Units

In order to detect transcriptional units, a mathematical object called *Genomic & Transcriptomic Segment (GTSegment)* was defined. This object results from a combination of (i) a co-expression network built from the above-mentioned transcriptomic data and (ii) the *Z. galactanivorans* genome organization. First, a co-expression network was built as follows. Genes, and by extension intergenic regions, with the same transcriptomic behavior were identified by computing for each pair of genes (g, g') the Spearman correlation $c_{(g, g')}$ and its related p -value $p_{(g, g')}$. Two genes have a similar expression when $c_{(g, g')} \geq 0.8$ and $p_{(g, g')} \leq 10^{-9}$ (after Bonferroni correction for multiple tests). For the sake of validation, each significant

correlation score was confirmed if identified by complementary *Mutual Information Coefficient* (MIC) ($cv = 0.6$) using MINE java programming (Reshef et al., 2011). We then built the co-expression network where nodes are the set of genes from *Z. galactanivorans* and edges connect two genes g and g' when g and g' have a similar expression. The co-expression network we obtained was composed of 3,912 nodes and 51,083 edges. For the sake of illustration, Supplementary Figure 2 represents the network via the use of a Hu algorithm, which groups highly interconnected nodes (Hu, 2005). In a second step, the *Z. galactanivorans* genome organization was modeled as a single circular sequence of genes, called G . The gene order in G is defined by the position of genes (and by extension intergenic regions) in the genome. A GTSegment S is then a sub-sequence of G where (i) S must contain at least 2 genes and at most 50 genes and (ii) a path in the co-expression network must exist between extremities of S only by passing through other genes from S . The quality of a GTSegment S is evaluated by its density $d_g(S)$, which describes the proportion of genes in the GTSegment that behave similarly. Its formal definition is the following:

$$d_g(S) = \frac{|R(S)|}{|S|} \quad (1)$$

where $R(S)$ is the set of genes such as, each gene g'' in $R(S)$ is a gene from S for which a path exists in the co-expression network from the extremities of S to g'' that only passes through other genes from S . GTSegments with density ≥ 0.6 are known to be good candidates for transcription units (Bordron et al., 2016).

As many GTSegments can be included within larger ones, we selected representative GTSegments, called *dominant GTSegments*. Given two GTSegments A and B , B dominates A if $R(A)$ is a subset of $R(B)$ (i.e., the part of the co-expression network involved in A is also involved in B), and $d_g(A) < d_g(B)$ (i.e., the proportion of genes from the co-expression network involved in B is larger than the one involved in A). A dominant GTSegment is then a GTSegment that is not dominated by any other GTSegment.

The Python package developed to compute GTSegments is freely available at <https://pypi.python.org/pypi/GTsegments>

RESULTS AND DISCUSSION

Utilization of Algal Polysaccharides Involves Genome-Wide Transcriptomic Shifts

To investigate transcriptomic changes linked to algal biomass utilization, *Z. galactanivorans* was grown in minimal medium with algal polysaccharides (alginate, laminarin, agar, porphyran, ι - or κ -carrageenan) or glucose used as a reference monosaccharide. These polysaccharides represent a large chemical diversity. Alginate is a polymer of 1,4-linked β -D-mannuronic acid and α -L-guluronic acid found in the cell wall of brown macroalgae. Laminarins are β -1,3 glucans with occasional β -1,6 branches used as carbon storage

compounds in brown macroalgae and in microalgae such as, diatoms and oomycetes. Agar, porphyran, and carrageenan are sulfated galactans abundant in the cell wall of red macroalgae. They consist of a linear backbone of galactose residues linked by alternating β -1,4 and α -1,3 glycosidic bonds. The main difference lies in the configuration of the α -linked galactose units, which are in the L configuration in agar and porphyran and in the D configuration in carrageenans, and defines two groups of red macroalgae known as agarophytes and carrageenophytes. Agar and porphyran differ by the presence of a 3,6-anhydro bridge or the sulfurylation on the O6 in the α -L-galactose unit, respectively. ι - and κ -carrageenans both feature sulfate esters on the O4 of the β -linked D-galactose residues and 3,6-anhydro-bridges in the α -D-galactose residues, but differ in the presence of additional sulfate esters on the O2 of the α -D-galactose residues for ι -carrageenan. *Z. galactanivorans* grew best with glucose, laminarin, and porphyran, reaching cell density of 1 after 48 h when samples were collected for gene expression profiling (Supplementary Figure 3). In comparison, cell density only reached ca. 0.4 with alginate and agar, and ca. 0.2 with κ - and ι -carrageenans. This could reflect differences in substrate accessibility or the presence of refractory fractions in larger polysaccharides.

A total of 3,823 genes and 1,074 intergenic regions were detected as significantly expressed in at least one sample ($p < 0.01$). NMDS of expression data showed an overall consistency of the transcriptomic profiles obtained from triplicate cultures. Despite the difference in final cell density between samples, the procedure clearly separated three groups of profiles based on the origin of the substrates, namely (i) glucose, (ii) polysaccharides from brown algae, and (iii) polysaccharides from red algae (Figure 1). This shows that *Z. galactanivorans* transcriptomes share common features when cells use polysaccharides from the same algal phylum (*Phaeophyceae* vs. *Rhodophyceae*). The similarity in the transcriptomes of cells growing on red algal substrates might partly reflect the fact that agar, porphyran, and carrageenans are polymers containing D-galactose residues. Nonetheless, these polysaccharides still feature large structural differences. Notably, agars and porphyrans also contain L-galactose residues that are absent from carrageenans. The sulfation patterns of these polysaccharides are also very different. The similarity in these transcriptomes is thus unlikely to be explained only by the presence of D-galactose in red algal sulfated galactans. In the case of the brown algal polysaccharides, laminarin (glucose polymer) induced a transcriptomic profile that clusters tightly with that obtained on alginate (polymer of mannuronic and guluronic acid) and away from that obtained on glucose. Therefore, our results seem to indicate that common regulations occur in response to different substrates from the same algal phylum, triggering the activation of specific transcriptomic programs dedicated to the degradation of either brown or red algal biomass.

Differential expression analysis identified a set of 748 genes that were significantly up- or down-regulated ($\text{FWER} \leq 0.05$, $|\log_2\text{FC}| > 2$) with at least one polysaccharide compared to

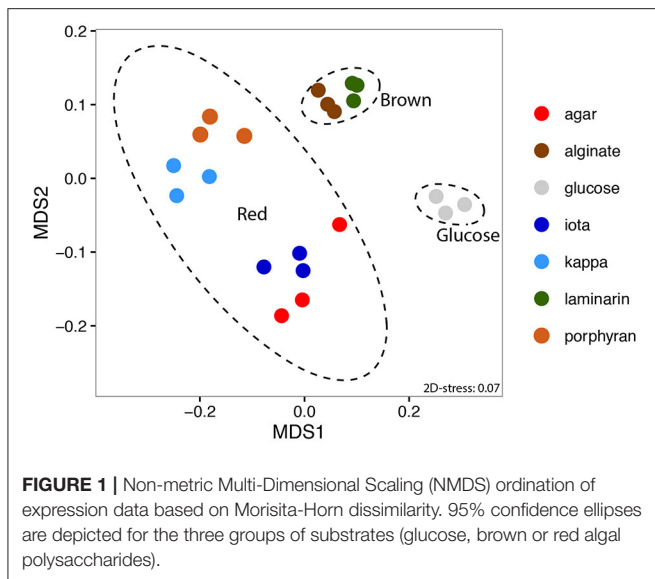


TABLE 1 | Global regulation of transcriptome depending on carbon source.

Carbon source	Up-regulated ^a	Down-regulated ^a	% regulated ^b
Alginate	63 genes (2.0–4.3)	128 genes (2.0–5.1)	4.26%
Laminarin	102 genes (2.0–4.6)	99 genes (2.0–5.1)	4.48%
Agar	106 genes (2.0–5.6)	91 genes (2.0–5.1)	4.40%
Porphyran	47 genes (2.0–4.9)	308 genes (2.0–4.6)	7.92%
κ -Carrageenan	65 genes (2.0–5.3)	311 genes (2.0–5.1)	8.40%
ι -Carrageenan	91 genes (2.0–4.7)	111 genes (2.0–5.1)	4.51%

^aDifferentially expressed genes compared to glucose condition (FWER <0.05, $|\log_2FC| > 2$). The range of $|\log_2FC|$ is given in brackets.

^bProportion of differentially expressed genes among all genes included on the chip.

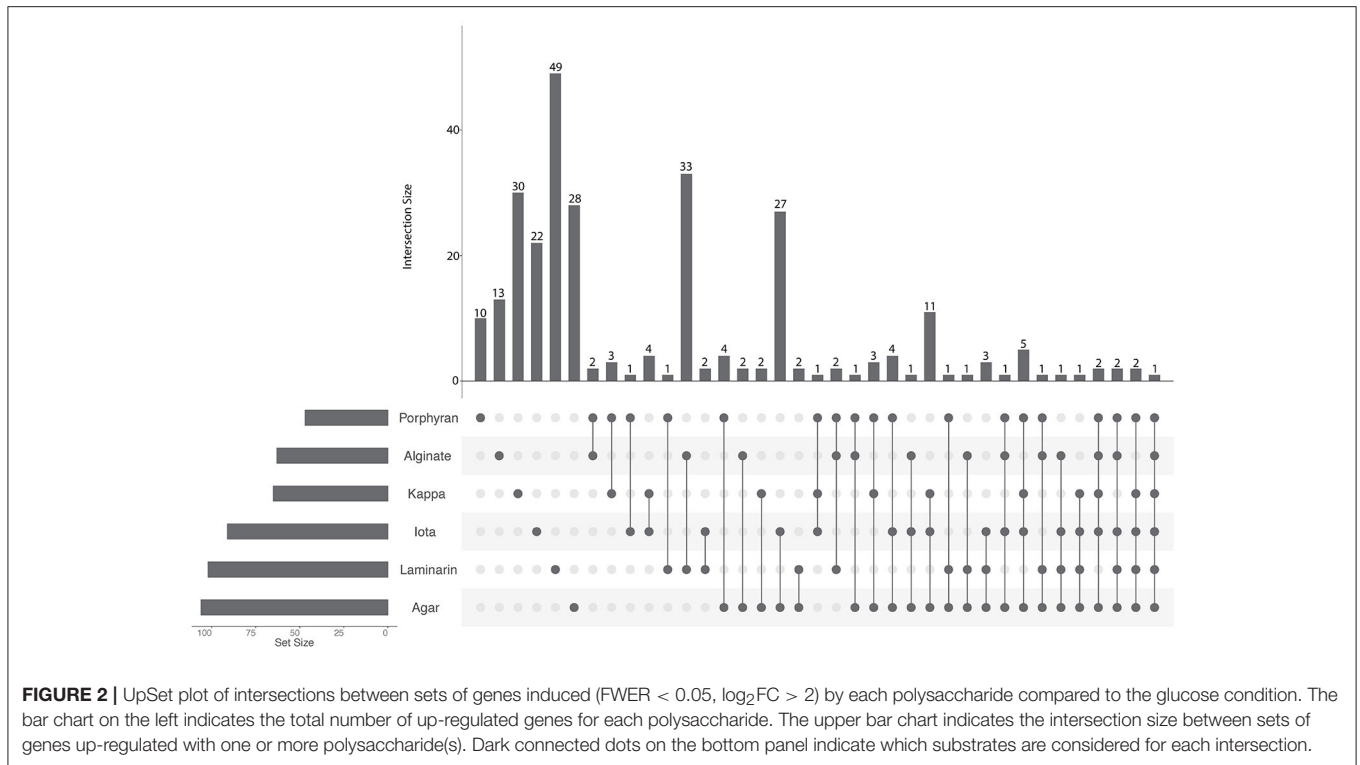
glucose used as a control condition (Supplementary Table 2). The proportion of differentially expressed genes varied from 4.3 to 8.4% depending on the carbon source (Table 1). The number of differentially expressed genes was about twice greater with porphyran and κ -carrageenan than with alginate, laminarin, agar, and ι -carrageenan. This was mainly due to a higher number of down-regulated genes. These additional down-regulated genes (Supplementary Table 3) were involved in various cellular functions, including transcription (e.g., RNA polymerase), translation (e.g., ribosomal proteins), energy generation (e.g., ATP synthase), and glucose utilization (e.g., 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, fructose bisphosphate aldolase, glucose-6-phosphate isomerase, glucose-6-phosphate-1-dehydrogenase). Unlike the four other polysaccharides, porphyran, and κ -carrageenan were not purchased commercially but rather purified directly from red macroalgae without acid or alkaline treatment. These extracted polysaccharides therefore likely better represent the complex structure of natural sulfated galactans found in red algal cell walls, such as, the presence of various substituents (sulfate, methyl, and pyruvate groups) that might be partially lost during the commercial preparation of polysaccharides. This could explain the larger effect on gene expression.

To validate the results, microarray data were compared with previously available RT-qPCR data obtained from independent experiments (Supplementary Figure 4). We considered the expression of 10 genes in the presence of laminarin, agar, porphyran, and alginate (Thomas et al., 2011a), as well as an additional set of 21 genes for alginate (Thomas et al., 2012). There was a strong positive correlation between \log_2 fold changes obtained previously by RT-qPCR and the microarray results [Pearson $r = 0.823$, $t_{(59)} = 11.13$, $p < 0.001$], confirming the robustness of the present data.

To identify substrate-specific or shared regulations, we compared the sets of up-regulated genes obtained with each polysaccharide (Figure 2, Supplementary Table 4). Half of the up-regulated genes (152 out of 279) responded specifically to the presence of one polysaccharide. Among the remaining 127 genes that were up-regulated by at least two different substrates, the highest intersection size was found between the transcriptomes of cells growing with alginate or laminarin (33 genes, representing 52 and 32% of the total number of up-regulated genes with alginate or laminarin, respectively). Similarly, intersection sizes were high for genes responding to agar and ι -carrageenan (27 genes), agar, ι -carrageenan, and κ -carrageenan (11 genes), or the four red algal polysaccharides (5 genes). By contrast, the intersection size for groups up-regulated both with brown and red algal polysaccharides did not exceed three genes (Figure 2). This corroborates the hypothesis that polysaccharides from the same algal phylum trigger shared regulations.

Expression Profiling Allows the Detection of Transcription Units

The global transcription shift induced by the different tested conditions offered an opportunity to investigate transcription units in *Z. galactanivorans*. In prokaryotes, many genes are organized into operons, defined as multiple open reading frames transcribed from the same promoter to a single mRNA transcript. Operons allow co-regulation at the transcriptional level for genes that often act as part of the same pathway. These structures can be predicted based solely on sequence information, using distances between adjacent genes, presence of promoters and terminators, and conservation in other genomes (Ermolaeva et al., 2001). Such a sequence-based search applied to the *Z. galactanivorans* genome predicted 1,046 operons containing at least two genes, as reported on the Database of Prokaryotic Operons DOOR 2.0 (Mao et al., 2014). However, these predictions need to be validated experimentally. The microarray dataset was therefore used to infer transcription units (GTSegments), defined as groups of proximal genomic elements displaying a coordinated expression in the tested conditions. The computation of dominant GTSegments in *Z. galactanivorans* using a custom script produced 278 transcription units of more than two and up to 17 genes (Figure 3, also available as high-resolution file in Supplementary Figure 5; Supplementary Table 5). These transcription units represented 192 unique regions distributed over the entire genome. A large proportion of the 1,046 total operons predicted by DOOR



was not detected from the co-expression network. This does not necessarily invalidate the DOOR predictions, but likely suggests that the set of tested growth conditions did not induce sufficient perturbations of the gene expression in all predicted operons to produce a detectable signal (Sabatti et al., 2002). Our experimental design, focused on the utilization of different algal carbon sources, is expected to better detect transcription units in the *Z. galactanivorans* genome for which the expression changes in the presence of polysaccharides.

A minority of 20% of the detected transcription units (55 GTSegments) featured genes transcribed in opposite directions (Supplementary Table 5), which necessarily belong to different operons. This co-expression of adjacent genes coded on opposite strands might arise from transcriptional coupling of closely spaced divergent promoters (Beck and Warren, 1988). By contrast, 80% of the detected transcription units (223 GTSegments) comprised genes transcribed all in the same direction and are therefore compatible with an operon structure. Of these, 136 transcription units were in agreement with the sequence-based DOOR predictions (Figure 3), providing empirical support for operon organization. A number of transcriptional units were represented by several GTSegments of different lengths, which could reflect the existence of secondary promoters within the operons, transcriptional attenuation, or differential degradation of mRNA. It is worth noting that this experimental approach without *a priori* succeeded in detecting transcription units known to be transcribed as genuine operons in other prokaryotic organisms. This includes operons involved in the biosynthesis of ribosomal proteins (Regions #25 and #93, Supplementary Table 5), tryptophan (Region #26),

arginine (Region #47), an F1F0 ATPase complex (Region #65), and a PUL dedicated to starch degradation (Region #15). Furthermore, novel operon-like structures potentially involved in algal polysaccharide utilization were detected (detailed below).

Polysaccharide-Responsive Regulons Provide New Candidate Genes Putatively Involved in Algal Substrate Recognition, Transport, and Degradation

Among the 279 genes up-regulated with at least one of the polysaccharides (FWER ≤ 0.05, log₂FC > 2), a large fraction (120 genes, i.e., 43%) encoded hypothetical proteins that lacked functional annotation. This number is significantly higher than what would be expected by chance, given that the microarray targets 1,429 genes encoding hypothetical proteins out of the 4,482 genes represented (Chi-squared test, $\chi^2 = 15.9$; $p < 0.001$). This enrichment of genes of unknown function within the polysaccharide-responsive regulons highlights the current lack of knowledge on catabolic pathways dedicated to marine polysaccharides, compared to glucose utilization for which the degradation machinery has been extensively studied. Furthermore, genes encoding proteins involved in transcriptional regulation were also significantly enriched among the up-regulated genes (Chi-squared test, $\chi^2 = 15.0$; $p < 0.001$). A total of 35 genes belonging to this category were up-regulated in at least one condition (Figure 4). This includes genes annotated as transcriptional regulators of the AraC, AsnC/Lrp, BlaI, HxlR, LysR, MarR,

