Microbiology (2004), **150**, 3571–3590 DOI 10.1099/mic.0.27041-0

Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria

Dmitry A. Rodionov,[1] Mikhail S. Gelfand[1,2] and Nicole Hugouvieux-Cotte-Pattat[3]

Correspondence
Dmitry Rodionov
rodionov@genetika.ru

[1]State Scientific Centre GosNIIGenetika, Moscow, 117545, Russia

[2]Institute for Problems of Information Transmission, Russian Academy of Sciences, Bolshoy Karetny per. 19, Moscow GSP-4, 127994, Russia

[3]Unité de Microbiologie et Génétique – Composante INSA, UMR CNRS-INSA-UCB 5122, bat Lwoff, 10 rue Dubois, Domaine Scientifique de la Doua, 69622 Villeurbanne Cedex, France

In the plant-pathogenic enterobacterium *Erwinia chrysanthemi*, almost all known genes involved in pectin catabolism are controlled by the transcriptional regulator KdgR. In this study, the comparative genomics approach was used to analyse the KdgR regulon in completely sequenced genomes of eight enterobacteria, including *Erw. chrysanthemi*, and two *Vibrio* species. Application of a signal recognition procedure complemented by operon structure and protein sequence analysis allowed identification of new candidate genes of the KdgR regulon. Most of these genes were found to be controlled by the cAMP-receptor protein, a global regulator of catabolic genes. At the next step, regulation of these genes in *Erw. chrysanthemi* was experimentally verified using *in vivo* transcriptional fusions and an attempt was made to clarify the functional role of the predicted genes in pectin catabolism. Interestingly, it was found that the KdgR protein, previously known as a repressor, positively regulates expression of two new members of the regulon, phosphoenolpyruvate synthase gene *ppsA* and an adjacent gene, *ydiA*, of unknown function. Other predicted regulon members, namely *chmX*, *dhfX*, *gntB*, *pykF*, *spiX*, *sotA*, *tpfX*, *yeeO* and *yjgK*, were found to be subject to classical negative regulation by KdgR. Possible roles of newly identified members of the *Erw. chrysanthemi* KdgR regulon, *chmX*, *dhfX*, *gntDBMNAC*, *spiX*, *tpfX*, *ydiA*, *yeeO*, *ygjV* and *yjgK*, in pectin catabolism are discussed. Finally, complete reconstruction of the KdgR regulons in various gamma-proteobacteria yielded a metabolic map reflecting a globally conserved pathway for the catabolism of pectin and its derivatives with variability in transport and enzymic capabilities among species. In particular, possible non-orthologous substitutes of isomerase KduI and a new oligogalacturonide transporter in the *Vibrio* species were detected.

Received 19 January 2004
Revised 25 June 2004
Accepted 11 August 2004

INTRODUCTION

Pectin is one of the signals detected by *Erw. chrysanthemi* that indicates the presence of plant tissues, triggering induction of pectinolysis and possibly of other associated virulence factors. Investigation of gene expression demonstrated that transcription of all genes involved in pectin catabolism is induced in the presence of pectin or its derivatives, such as polygalacturonate (PGA) and galacturonate (GA) (Hugouvieux-Cotte-Pattat *et al.*, 1992). This included genes encoding various types of pectinases, i.e. pectin acetylesterases (*paeX*, *paeY*), pectin methylesterases (*pemA*, *pemB*), pectate lyases (*pelA*, *pelB*, *pelC*, *pelD*, *pelE*, *pelI*, *pelL*, *pelW*, *pelZ*, *pelX*) and polygalacturonases (*pehN*, *pehV*, *pehW*, *pehX*), and also proteins necessary for secretion of pectinases (the *outC–M* operon), transporters of pectic oligomers (*kdgM*, *togT*, *togMNAB*) and intracellular enzymes involved in the cleavage of dimers (*ogl*) and the catabolism of unsaturated monomers (*kduI*, *kduD*, *kdgK*) (for a review, see Robert-Baudouy *et al.*, 2000).

Induction of several genes in the presence of pectin indicated a co-ordinated regulation of these genes. KdgR, which belongs to the IclR family of transcriptional regulators, has been characterized as being responsible for this regulation (Nasser *et al.*, 1992). In a *kdgR* mutant, the expression of all these genes, except *pelL*, increased, indicating that they

Abbreviations: CRP, cAMP-receptor protein; DK-I, 5-keto-4-deoxyuronate; DK-II, 2,5-diketo-3-deoxygluconate; GA, galacturonate; KDG, 2-keto-3-deoxygluconate; MCP, methyl-accepting chemotaxis protein; OGA, oligogalacturonate; PGA, polygalacturonate.

are repressed *in vivo* by KdgR. *In vitro* analysis demonstrated that KdgR directly interacts with the promoter regions of the *in vivo*-controlled genes/operons (Nasser *et al.*, 1994). KdgR-binding sites usually overlap with or are close to the promoters. These observations suggest that the KdgR protein and the RNA polymerase compete for adjacent binding sites on DNA, explaining how KdgR binding prevents gene expression. Physiological and biochemical studies indicated that an intermediate of pectin catabolism, 2-keto-3-deoxygluconate (KDG), is the main inducing molecule which interacts *in vivo* and *in vitro* with KdgR, provoking dissociation of KdgR from its operators. *In vivo* data indicate that two other unsaturated monomers formed during pectin catabolism, 5-keto-4-deoxyuronate (DK-I) and 2,5-diketo-3-deoxygluconate (DK-II), are also able to act as inducers by interaction with KdgR. From current data, the KdgR repressor directly controls at least 13 operons that constitute the KdgR regulon (Hugouvieux-Cotte-Pattat *et al.*, 1996). A genetic screen using *lacZ* transcriptional fusions indicated that as much as 1 % of the *Erw. chrysanthemi* genes (about 50 genes) could be induced in the presence of pectin (Hugouvieux-Cotte-Pattat & Robert-Baudouy, 1989). Identification of some of these genes has confirmed the presence of known pectinase genes, but has also revealed new pectin-inducible loci, such as the recently identified *rhiTN* operon involved in the catabolism of a pectin-related plant polysaccharide, rhamnogalacturonan (Hugouvieux-Cotte-Pattat, 2004). These observations show that KdgR has a wide range of targets and its role may not be restricted to pectinolysis.

KdgR homologues were also identified in other plant-pathogenic enterobacteria, *Erwinia carotovora* subsp. *carotovora*, *Erw. carotovora* subsp. *atroseptica* and *Erw. amylovora* (Liu *et al.*, 1999; Thomson *et al.*, 1999). In animal-related enterobacteria, such as *Escherichia coli*, the action of KdgR was found to be restricted to the control of expression of genes involved in the catabolism of KDG (*kdgK*, *kdgA* and *kdgT*). These results indicate that KdgR is a regulatory protein conserved in the *Enterobacteriaceae*. Moreover, the KdgR proteins appeared to be functionally interchangeable between species. For instance, KdgR of *E. coli* is able to repress *in vivo* transcription of the *Erw. chrysanthemi pelD* gene (James & Hugouvieux-Cotte-Pattat, 1996), whereas KdgR of *Erw. chrysanthemi* is able to bind *in vitro* to regulatory regions of pectinase genes from *Erw. carotovora* and vice versa (Thomson *et al.*, 1999). Thus, conservation of the KdgR regulator is accompanied by conservation of its specific binding signals.

Comparative genomics is a powerful approach for the prediction of gene regulation and the annotation of the bacterial genome (Gelfand *et al.*, 2000; Gelfand, 2003). Previous *in silico* analysis of the KdgR regulon revealed several novel KdgR-regulated genes in gamma-proteobacteria (Rodionov *et al.*, 2000), such as the predicted oligogalacturonide transporter OgtABCD, which was confirmed in an independent experimental study to have the proposed function (renamed as TogMNAB) and to be regulated *in vivo* by KdgR in *Erw. chrysanthemi* (Hugouvieux-Cotte-Pattat *et al.*, 2001). Recent availability of many complete genomes of enterobacteria, including the two plant pathogens *Erw. carotovora* (http://www.sanger.ac.uk/Projects/E_carotovora) and *Erw. chrysanthemi* (http://www.tigr.org/tdb/mdb/mdbinprogress.html), provides an opportunity to perform more detailed comparative analysis of the KdgR regulon in a variety of bacteria. This has allowed us to identify a large number of new KdgR regulon members. A complete description of the KdgR regulon in enterobacteria and *Vibrio* species has revealed the main differences in the pectin and KDG degradation pathways in these bacteria. We took advantage of the genetic tools and knowledge obtained in *Erw. chrysanthemi* to validate some of the data resulting from this comparative analysis.

METHODS

Datasets and genomic sequences. Complete genomic sequences of *E. coli* K-12 and CFT073, *Salmonella typhimurium*, *Yersinia pestis*, *Vibrio vulnificus* and *Vibrio parahaemolyticus* with their annotation files were obtained from GenBank (Benson *et al.*, 2000). Unannotated contig sequences of *Erwinia chrysanthemi* and *Klebsiella pneumoniae* were downloaded from the websites of the Institute for Genomic Research (www.tigr.org) and the Washington University Consortium (www.genome.wustl.edu), respectively. The *Erwinia carotovora* subsp. *atroseptica* and *Yersinia enterocolitica* complete sequence data were produced by the respective sequencing groups at the Sanger Institute and were obtained from ftp://ftp.sanger.ac.uk/pub/.

Identification of DNA-binding motifs. All previously characterized KdgR-binding sites in *Erw. chrysanthemi* were collected from the literature (Hugouvieux-Cotte-Pattat *et al.*, 1996). The KdgR search profile was constructed using an alignment of these known sites. Positional nucleotide weights in this profile were derived using the following formula:

$$W(b, k) = \log[N(b, k) + 0{\cdot}5] - 0{\cdot}25 \sum_{i=A,C,G,T} \log[N(i, k) + 0{\cdot}5]$$

where $N(b, k)$ is the count of nucleotide b in position k in the training sample of aligned sites. The consensus for the KdgR sites is the 21 bp sequence WAWTRAAAYRnYRTTTYAWTW. The score of a candidate site was calculated as the sum of the positional nucleotide weights:

$$Z(b_1, b_L) = \sum_{k=1,L} W(b_k, k)$$

where $L = 21$ is the length of the KdgR signal. The site score defined by this formula is linearly related to the discrimination energy and can be used to assess the significance of individual sites (Mironov *et al.*, 1999). Then, each genome was scanned with the KdgR profile, and genes with candidate regulatory sites in upstream regions (normally in positions -300 to $+50$ relative to the translation start) were selected. The cut-off score for putative KdgR-binding sites in closely related enterobacteria was defined as a lowest score within the training set (5·20). In the case of more distant *Vibrio* species, the KdgR search profile was derived from a set of upstream regions of orthologous KDG genes using the SignalX program (Mironov *et al.*, 2000). The cut-off score for candidate KdgR sites in these genomes was 5·0. To account for possible operon structures, the resulting set of candidate regulon members was supplemented by genes that are likely to be co-transcribed with genes preceded by candidate sites (with an intergenic distance less than 100 bp).

The recognition profiles for the catabolic regulatory proteins CRP (cAMP receptor protein) and FruR were constructed using the same procedure and training sets of 70 known CRP-binding sites and 12 known FruR-binding sites were collected from the literature (data not shown). Consensus sequences for the CRP and FruR sites are WWWTGTGATNNNNATCACAWWW and GCTGAAWCGWTTC-AGC, respectively. The search profile for RhaS sites was kindly provided by O. Laikova (Gelfand & Laikova, 2003).

Other computer programs. The signal recognition procedure and the Smith–Waterman alignment of protein sequences were performed using the Genome Explorer program (Mironov *et al.*, 2000). Orthologous genes in studied gamma-proteobacteria were identified by the bidirectional best hits criterion (Tatusov *et al.*, 2000). Additional protein sequence comparisons and search of distant homologues in protein databases were performed using gapped BLASTP and PSI-BLAST programs (Altschul *et al.*, 1997). If necessary, orthologous or paralogous relationships of proteins were confirmed by construction of phylogenetic trees. The phylogenetic trees were constructed by the maximum-likelihood method implemented in PHYLIP (Felsenstein, 1981). Multiple protein sequence alignments were constructed by CLUSTALX (Thompson *et al.*, 1997). Potential transmembrane segments and signal peptide cleavage sites were predicted using the TMpred (www.ch.embnet.org/software/TMPRED_form.html) and SignalP (www.cbs.dtu.dk/services/SignalP/) servers, respectively (Hofmann & Stoffel, 1993; Nielsen *et al.*, 1997).

Strains, media and growth conditions. The bacterial strains of *Erw. chrysanthemi* and the plasmids used in this study are listed in Table 1. The Phi-EC2 generalized transducing phage was used for transduction (Resibois *et al.*, 1984). *Erw. chrysanthemi* cells were grown at 30 °C in M63 medium (Miller, 1972). Carbon sources, namely glycerol, GA and PGA, were added at 2 g l^{-1}. *E. coli* cells were grown at 37 °C in LB medium (Miller, 1972). The media were solidified with agar (15 g l^{-1}). When required, antibiotics were added at the following concentrations: kanamycin (Km), 20 μg ml^{-1}; ampicillin, 50 μg ml^{-1}; chloramphenicol, 20 μg ml^{-1}.

Chemotaxis was measured by determining the size of haloes observed on semi-solid agar plates containing 4 g agar l^{-1} in M63 medium (Hugouvieux-Cotte-Pattat *et al.*, 2001). Oligogalacturonides were prepared by degradation of PGA with pectate lyases (Hugouvieux-Cotte-Pattat *et al.*, 2001). For these experiments, 0·2 mM glycerol was added as the carbon source and attractants were used at a final concentration of 1 mM. The diameters of the chemotactic rings were measured after incubation for 24 h at 30 °C.

Recombinant DNA techniques. Preparation of plasmid or chromosomal DNA, restriction digestions, ligations, DNA electrophoresis and transformations were carried out as described by Sambrook *et al.* (1989).

PCR primers were designed (24- to 28-mers, Table 1) to clone 0·7–1·3 kb chromosomal DNA containing the entire gene or its 5′ end; restriction sites were added at each end to determine the orientation of the DNA insertion in the vector (*Bam*HI or *Bgl*II at the 5′ end and *Xba*I at the 3′ end). Strain 3937 chromosomal DNA was used as the template. The PCR products were purified (QIAquick PCR purification kit; Qiagen) and directly ligated to the pGEM^{R}-T vector (Promega) which has a protruding T nucleotide at each 3′ end.

Genetic fusions were constructed on the cloned genes, by insertion of *uidA*-Km cassettes (Bardonnet & Blanco, 1992) into a restriction site situated inside the corresponding ORF (Table 1). The orientation of the *uidA*-Km cassette was determined by restriction analysis. Only plasmids in which *uidA* and the mutated gene have the same transcriptional direction were retained. Plasmids bearing the *uidA*-Km insertion were then introduced into *Erw. chrysanthemi* cells by electroporation. The insertions were integrated into the *Erw. chrysanthemi* chromosome by marker exchange recombination after successive cultures in low phosphate medium supplemented with Km (Roeder & Collmer, 1985). After verification of the correct recombination of the *uidA*-Km insertions by PCR, β-glucuronidase activity was measured to estimate the expression of the fused gene. The degradation of *p*-nitrophenyl-β-D-glucuronide into *p*-nitrophenol, was followed at 405 nm. Specific activity is expressed as nmol products liberated min^{-1} (mg bacterial dry wt)$^{-1}$.

RESULTS AND DISCUSSION

Characterization of the KdgR regulons in enterobacteria and *Vibrio* species

Existence of the KdgR orthologue is a prerequisite to the comparative analysis of the KdgR regulons in bacteria. Based on the phylogenetic tree of the IclR family homologues from various bacteria (Table 2 and data not shown), we identified KdgR in all studied enterobacteria: two *Erwinia* species, *Erw. chrysanthemi* (*ER*) and *Erw. carotovora* (*EO*), two *Yersinia* species, *Y. pestis* (*YP*) and *Y. enterocolitica* (*YE*), *K. pneumoniae* (*KP*), *S. typhimurium* (*ST*) and *E. coli* (*EC*), and in two *Vibrio* species, *V. vulnificus* (*VV*) and *V. parahaemolyticus* (*VP*). For *EC*, strain CFT073 was also considered since it contains more KdgR-controlled genes than strain K-12. A high degree of sequence conservation in the KdgR proteins implies conservation of KdgR-binding signals in all considered species of enterobacteria. Known *ER* KdgR-binding sites were collected from previous studies and comprised the training set for a 21-bp recognition profile (Fig. 1a). Then, the KdgR profile was used to search for new candidate KdgR-binding sites in the genomes of *ER* and other enterobacteria. Table 3 lists both previously known and newly identified KdgR-binding sites.

Since KdgR orthologues detected in the *Vibrio* species are less similar to KdgR from enterobacteria (Table 2), we tried to construct a more specific profile of the KdgR-binding sites in *VV* and *VP*. Towards this aim, we selected the regions upstream of the orthologues of the *ER* KdgR-regulated genes in the genomes of *VV* and *VP*. As a result, a common 21-bp palindromic signal highly similar to the KdgR site from enterobacteria was obtained (Fig. 1b) and was used for identification of new members of the KdgR regulon in both *Vibrio* species (Table 3). In contrast to the *VV* and *VP* genomes, orthologues of *kdgR* and of genes involved in pectin/KDG catabolism were not detected in the genome of *Vibrio cholerae*, arguing for possible recent loss of the complete KdgR regulon in this highly pathogenic bacterium.

Almost all previously known KdgR-regulated genes in *ER* are subject to catabolic repression by glucose moderated through the CRP (Reverchon *et al.*, 1997). To test whether this global CRP regulation is conserved for other members of the KdgR regulon, we scanned all studied genomes with the CRP profile. The online version of this paper (at http://mic.sgmjournals.org) contains a supplementary table showing the list of all candidate CRP sites found upstream of the

Table 1. Bacterial strains, plasmids and oligonucleotides used in this study

Strain/plasmid	Genotype/phenotype	Reference/origin
***Erw. chrysanthemi* strains**		
3937	Wild-type	Laboratory collection
A350	*lmrT*^{c} *lacZ2*	Hugouvieux-Cotte-Pattat *et al.* (1989)
A576	*lmrT*^{c} *lacZ2 arg kdgK*	Laboratory collection
A1077	*lmrT*^{c} *lacZ2 kdgR*::Cm	Laboratory collection
A4260	*lmrT*^{c} *lacZ2 ydiA*::*uidA*, Km	This work
A4261	*lmrT*^{c} *lacZ2 ppsA*::*uidA*, Km	This work
A4302	*lmrT*^{c} *lacZ2 spiX*::*uidA*, Km	This work
A4303	*lmrT*^{c} *lacZ2 chmX*::*uidA*, Km	This work
A4304	*lmrT*^{c} *lacZ2 yjgK*::*uidA*, Km	This work
A4307	*lmrT*^{c} *lacZ2 sotA*::*uidA*, Km	This work
A4310	*lmrT*^{c} *lacZ2 dhfX*::*uidA*, Km	This work
A4370	*lmrT*^{c} *lacZ2 gntB*::*uidA*, Km	This work
A4371	*lmrT*^{c} *lacZ2 tpfX*::*uidA*, Km	This work
A4372	*lmrT*^{c} *lacZ2 yeeO*::*uidA*, Km	This work
A4373	*lmrT*^{c} *lacZ2 pykF* ::*uidA*, Km	This work
Plasmids		
pGEM-T	Vector for direct cloning of PCR product, Ap^R	Promega
pI2877	pGEM-T derivative with a 1 kb PCR fragment, '*ppsA*'$^+$	This work
pI2879	pGEM-T derivative with a 1·1 kb PCR fragment, *ydiA*′$^+$	This work
pI2884	pGEM-T derivative with a 1·1 kb PCR fragment, *yeeO*′$^+$	This work
pI2886	pGEM-T derivative with a 0·9 kb PCR fragment, *yjgK*$^+$	This work
pI2928	pGEM-T derivative with a 1·3 kb PCR fragment, *chmX*′$^+$	This work
pI2929	pGEM-T derivative with a 1 kb PCR fragment, *dhfX*′$^+$	This work
pI2931	pGEM-T derivative with a 1 kb PCR fragment, *gntD2*′$^+$	This work
pI2932	pGEM-T derivative with a 1 kb PCR fragment, *spiX*$^+$	This work
pI2933	pGEM-T derivative with a 1 kb PCR fragment, *gntD*′$^+$	This work
pI2934	pGEM-T derivative with a 1 kb PCR fragment, *sotA*′$^+$	This work
pI2983	pGEM-T derivative with a 1 kb PCR fragment, '*gntB*'$^+$	This work
pI2984	pGEM-T derivative with a 1·1 kb PCR fragment, *pyk*F′$^+$	This work
pI2985	pGEM-T derivative with a 1·2 kb PCR fragment, *tpfX*$^+$	This work
pI2986	pGEM-T derivative with a 1·3 kb PCR fragment, *ygjV*$^+$	This work
pI2878	pI2877 derivative with a *uidA*-Km cassette in the *Pst*I site, *ppsA*::*uidA*, Km^R	This work
pI2880	pI2879 derivative with a *uidA*-Km cassette in the *Pst*I site, *ydiA*::*uidA*, Km^R	This work
pI2887	pI2886 derivative with a *uidA*-Km cassette in the *Sal*I site, *yjgK*::*uidA*, Km^R	This work
pI2935	pI2932 derivative with a *uidA*-Km cassette in the *Sma*I site, *spiX*::*uidA*, Km^R	This work
pI2936	pI2928 derivative with a *uidA*-Km cassette in the *Eco*RI site, *chmX*::*uidA*, Km^R	This work
pI2943	pI2933 derivative with a *uidA*-Km cassette in the *Bam*HI site, *gntD*::*uidA*, Km^R	This work
pI2945	pI2934 derivative with a *uidA*-Km cassette in the *Pst*I site, *sotA*::*uidA*, Km^R	This work
pI2946	pI2929 derivative with a *uidA*-Km cassette in the *Sal*I site, *dhfX*::*uidA*, Km^R	This work
pI2948	pI2931 derivative with a *uidA*-Km cassette in the *Bam*HI site, *gntD2*::*uidA*, Km^R	This work
pI2990	pI2884 derivative with a *uidA*-Km cassette in the *Nco*I site, *yeeO*::*uidA*, Km^R	This work
pI2992	pI2985 derivative with a *uidA*-Km cassette in the *Nru*I site, *tpfX*::*uidA*, Km^R	This work

Table 1. cont.

Strain/plasmid	Genotype/phenotype	Reference/origin
pI3005	pI2983 derivative with a *uidA*-Km cassette in the *Eco*RV site, *gntB*::*uidA*, Km^R	This work
pI3006	pI2984 derivative with a *uidA*-Km cassette in the *Sal*I site, *pykF*::*uidA*, Km^R	This work
pI3007	pI2986 derivative with a *uidA*-Km cassette in the *Sal*I site, *ygjV*::*uidA*, Km^R	This work
Oligonucleotides*		
chmX	GCGGATCCTGACCGTTTCTGTTGACC	CGTCTAGATTGGCGTTATCCGAGTTC
dhfX	GCGGATCCTATCCGGCACTTGTTGCC	CGTCTAGATGCAGCATGTAAAAGGAG
gntB	GCGGATCCTTCAATCTGGTCGATAACGCG	CGTCTAGACCAGCATGTCGAACTTGC
gntD	GCAGATCTAGTGATAATGCTCACAAGGC	CGTCTAGACGCCATGCTCTGCTCTTC
gntD2	GCAGATCTACCAGACCGGTTCAGACAGC	CGTCTAGATGACGTCGATGTCCAGGTTC
ppsA	GCGGATCCATGACATTAGCGAAATGCG	CGTCTAGAACCGATAGAGAAACCGTCG
pykF	GCGGATCCTCTCGCAGTCGCAAACGTATTG	CGTCTAGACCGCTTCCAGCGGGTATTTGC
sotA	GCGGATCCTCTCTTGCTGAACTTG	CGTCTAGATGCCGTTGCGGTCCAC
spiX	GCGGATCCGCCGAATTGTTGTCTGG	CGTCTAGATGTAACGGATCTGATTACGG
tpfX	GCGGATCCGATCTGGGAAGCGATCGTG	CGTCTAGAAATTTGCTGGTGTCTACGTC
ydiA	GCGGATCCATACAGTTAGGATGTGAAG	CGTCTAGATGTAACGGATCTGATTACGG
yeeO	GCGGATCCGCTATATCTGGTACATCAG	CGTCTAGAGAATCGTGGTGGTGAACG
ygjV	GCGGATCCGTGACCTTGTAAAGCGGATCG	CGTCTAGATGGCGTTGGCTAAGGCAGAC
yjgK	GCGGATCCTAGCGGATTGACCATGTGGC	CGTCTAGATAGGCGTGACGGCTTTCG

*Restriction sites are underlined.

KdgR-controlled genes. This analysis suggests that the majority of the pectin degradation and utilization genes in gamma-proteobacteria are under dual regulation by KdgR and CRP (Fig. 2). Moreover, the relative positions of the candidate CRP and KdgR-binding sites agree with the known antagonistic effect of CRP and KdgR on the expression of the pectinolytic genes in *ER* (Nasser *et al.*, 1997).

The operon structures of the KdgR-regulated genes and the location of potential KdgR and CRP sites are summarized in Fig. 2. The predicted KdgR regulons of two *Erwinia* species are particularly large, containing 31 operons for *ER* and 29 operons for *EO* (Fig. 2). These regulons include a variety of extracellular pectinolytic enzymes (Pel, Peh, Pem and Pae), a secretion system (Out), several porins (KdgM) and transport systems (TogMNAB, TogT, KdgT, etc.), as well as enzymes for the intracellular catabolism of dimers

Table 2. Percentage identity between the KdgR proteins from enterobacteria and *Vibrio* sp.

Genome	*ER*	*EO*	*YE*	*YP*	*KP*	*ST*	*EC*	*VP*	*VV*
ER	100								
EO	92	100							
YE	88	88	100						
YP	89	88	97	100					
KP	84	82	81	81	100				
ST	89	86	84	85	93	100			
EC	91	87	86	86	89	96	100		
VP	68	68	65	67	73	66	67	100	
VV	66	66	63	65	73	64	66	91	100

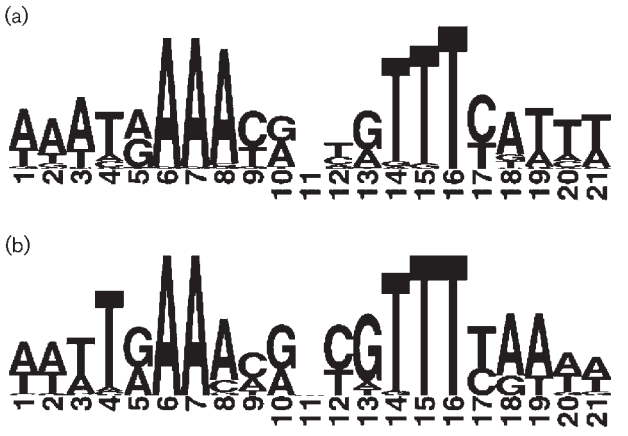

Fig. 1. Sequence logos for the KdgR-binding sites in enterobacteria and *Vibrio* spp. (a) The KdgR motif drawn from a training set of known *Erw. chrysanthemi* sites; (b) the most significant motif obtained by the signal determination procedure for *Vibrio* spp.

Table 3. Predicted KdgR regulon in enterobacteria and *Vibrio* spp.

Genes marked with an asterisk (*) were named in this study. Divergently located genes are separated by /. Lower case letters in the site sequences indicate positions that do not conform to the consensus. Site scores lower than 5·20 for enterobacteria and lower than 5·00 for *Vibrio* sp. are underlined and correspond to weak sites. The table contains all candidate KdgR sites with a score higher than the respective cut-offs, and also several weak sites that either are conserved in other species or precede pectin degradation genes. The last column represents the experimental data on regulation: S, *in vivo* and *in vitro* functional KdgR-binding sites; R, previously known *in vivo* regulation by KdgR; EC, regulation by KdgR confirmed by experiments in this study; EN, regulation by KdgR could not be confirmed.

Operons	Function	Site	Position	Score	Regulation
Erw. chrysanthemi					
kdgT	Transporter of KDG, DK-I and DK-II	AAAaGAAACATTGTTTCATTT	−229	6·07	S
kduI-kduD/kdgF	DK-I catabolism	AAATaAAACATTaTTTCATTT	−189	6·05	S
pelX	Exopectate lyase	AAAaGAAACAgTGTTTCATTT	−67	5·99	R
ygjV	?	AAATaAAACggcGTTTCATTa	−67	5·94	?
*tpfX**	?	tAtTGAAACgATaTTTCATTT	−56	5·92	EC
out	Pectinase secretion system	tAATGAAACggTGTTTtATTa	−144	5·83	S
togT	Oligogalacturonide transporter	AAATGAAAtAATGTTTtAaTa	−40	5·79	S
pelB	Pectate lyase	tAATGAAAtggcaTTTCAaTT	−177	5·72	S
*chmX**	~ MCP receptor	AAATaAAAtgTTGTTTtgTTT	−173	5·67	EC
		tAAaGAAACAAcaaTgCAaTT	−195	4·51	EC
kdgK	KDG kinase	AAATaAAACATcGTTTCATcg	−101	5·64	S
ogl	OGA lyase	AAATGAAACgTTGTTTCtaca	−111	5·63	S
		AAATGAAAgAATGTTTtATaa	−153	5·15	S
pehN	Polygalacturonase	AAATGAAACgTTGTTTtAcTT	−32	5·62	S
ppsA/ydiA	Phosphoenolpyruvate synthase/?	tAtTaAAACAcTaTTTCATTa	−220	5·59	EC+
pelE	Pectate lyase	AgATGAAAtggTaTTTCgTTT	0	5·59	S
pelA	Pectate lyase	AttTaAAACATcGTTTCATTa	−236	5·52	S
kdgN	OGA porin paralogue	AAATGAAAtAgctTTTCATTT	−125	5·58	R
pelC-pelZ	Pectate lyases	tAATGAAAttAcGTTTCAacT	−199	5·44	S
pelW-togMNAB	Oligogalacturonide catabolism and transport	AAtcaAAACAATGTTTCtaTT	−61	5·42	S
*gntD2**	?	AAATaAAACgcaaTTTtATTa	−57	5·39	?
yeeO	~ Multidrug efflux transporter	tttTaAAACATcGTTTCATcT	−216	5·35	EC
*spiX**	? ~ Sugar isomerases	AAATGAAAtAgTGTTTtAaaa	−55	5·34	EC
yjgK	?	AAtcaAAACATTGTTTCAgTT	−43	5·26	EC
*dhfX**	?	AAAaGgAACgcTGTTTtATTT	−35	5·25	EC
pelD-paeY-pemA	Pectate lyase, acetylesterase, methylesterase A	ggAcaAAAtggcGTTTCATTT	−82	5·21	S
pemB	Pectin methylesterase B	AAATGAAACgcaGgTTtATTT	−171	5·17	R
cheX13	~ MCP receptor	AgtTGAAgtgcTaTTTCATTT	−185	5·09	?
*gntDBMNAC**	?	AAATaAAACgTgGTTTtcTTT	−297	5·00	EC
		AAtTaAtACATgaTTTCtTTT	−275	4·61	EC
cheX21	~ MCP receptor	ttATGAAACtcTaTTTCATTc	−213	4·97	?
pecT	Regulator of pectate lyases	AAtTaAAACAggaTTTCAaaT	−478	4·84	EC
pehX	Polygalacturonase	ttATaAAACgTcGTTTCgaaa	−79	4·79	S
pykF	Pyruvate kinase I, fructose-stimulated	AtAcGgAACgTcGTTTCATTg	−275	4·75	EC
pir	Regulator of pectinolysis	AcATaAAAtgcTGTTTCATga	−15	4·76	EN
expI	Regulation of extracellular enzyme production	AAAgGAAAtAATaTgTCAacT	−137	4·65	EN
pehW	Polygalacturonase	ttATaAAACATaGTTTCtTaa	−87	4·64	S
pelI	Pectate lyase	tttcGAAACAgcGgTTtATTT	−139	4·62	S
sotA	Sugar efflux transporter	AAcTGAAACggacTTTCgTTT	−85	4·62	EC
rhiT-rhiN	Rhamnogalacturonide transport and catabolism	AAAcGAAcCgTTGgTTttaTT	−207	4·37	R
pehV	Polygalacturonase	ttATaAAACgTgaTTTCtTaa	−91	4·35	S

Table 3. cont.

Operons	Function	Site	Position	Score	Regulation
Erw. carotovora					
ygjV	?	AAATaAAACggcGTTTCATTT	−69	6·06	
kduI-kduD/kdgF	DK-I catabolism	AAATaAAACATTaTTTCATTT	−210	6·05	
*tpfX**	?	tAATaAAACATcGTTTCATTT	−63	6·03	
pelX	ExoPGA lyase	AAAaGAAACAgcGTTTCATTT	−67	5·98	
		AAATGAcACggTGTTTCATaa	−131	5·13	
*dhfX**	?	AAAaGAAACgATGTTTtATTT	−109	5·92	
ogl	OGA lyase	AAATGAAACATTGTTTCtaTa	−179	5·81	
rexZ	Regulator of exoenzyme production	tAATaAAACtATGTTTCATTT	−139	5·73	S
pehN	Polygalacturonase	tAtTaAAACgATaTTTCATTa	−36	5·66	
*gntDBMNAC**	?	AAATaAAACgcTaTTTttTTT	−174	5·59	
ppsA/ydiA	Phosphoenolpyruvate synthase/?	tAtTaAAACgccaTTTCATTa	−205	5·57	
*sghX**	? ∼ Glycosyl hydrolase	AAATaAAACcTcGTTTCtTTT	−67	5·44	
pel3-pelZ	Pectate lyases	AAAcGAAAtATaaTTTtATTT	−288	5·40	S
kdgK	KDG kinase	AAATaAAACATcGTTTCAaga	−121	5·35	
pelW-togMNAB	Oligogalacturonides transport and catabolism	AAtcaAAACAgTGTTTCtaTT	−62	5·34	
		AAtcaAAACAAcGTTcCgacT	−140	4·65	
togT	Oligogalacturonide transporter	AAcTGAAAtATTGTTTCATaT	−39	5·29	
		tAAaGAAAtgATagTTCAaTc	−61	4·65	
*rhiABC**	Predicted transporter for rhamnogalacturonides	AtATaAAACAgcaTTTCATaT	−240	5·22	
kdgM3-kdgM4-pelP	OGA porin paralogues, pectate lyase	AAATGAAACAATGcTTCcTTT	−348	5·18	
paeY-pemA	Pectin acetylesterase, methylesterase A	tAtTaAAAtAAcaTTTCAcTT	−214	5·14	
		tAATGAAttgTcaTTTCAcca	−192	4·53	
yeeO	∼ Multidrug efflux transporter	AAtaGAAACgTcaTTTtAaaa	−94	5·03	
pykF	Pyruvate kinase I, fructose-stimulated	AtAcGgAACgTcGTTTCAaTT	−257	4·95	
*spiX**	? ∼ Sugar isomerases	AAtTGAAAtAcaGTTTtAaaa	−55	4·89	
sotA	Sugar efflux transporter	tAATGAAACAAgaTTTCtgTT	−89	4·83	
pel-pehX-out*	Pectate lyase, polygalacturonase, secretion system	tAATGAAACcATGTTTtAcca	−290	4·74	
*cheX**	∼ MCP receptor	tAtTGAAACgTgaTTTCtaTc	−183	4·73	
yjgK	?	ttgTaAAACAggGTTTCATTT	−42	4·58	
rsmB	Regulatory RNA involved in pathogenesis	TGTTGAAAATAGGTTTCATTT	−67	4·56	R
*kdgX**	? Predicted KDG permease	gAtaaAAAagTTGTTTttTTT	−282	4·55	
rhiT-rhiN	Rhamnogalacturonides transport and catabolism	AgtTaAAAtgcctTTTtATaa	−160	4·49	
kdgM-paeX	OGA porin, pectate acetylesterase	AttTGAAACccTaTTTtAcTT	−154	4·45	
pel1	Pectate lyase	AgAcaAAcCtTaaTTTCATTT	−134	4·44	S
pel2	Pectate lyase	tAAaGgAAtATcGTTTCcTaT	−418	4·34	
pelI	Pectate lyase	AgAaaAAAtATccgTTtAaTT	−35	4·25	
pehA	Polygalacturonase	AAAgGAAgCgTaaTTcCtTTT	−66	3·93	R
Y. enterocolitica					
yjgK	?	tAATGAAACATcaTTTCATTT	−43	6·06	
*kdgX**	? Predicted KDG permease	tAAcGAAACAATGTTTCATTT	−42	5·95	
kduI-kduD/kdgF	DK-I catabolism	tAATaAAACATcaTTTCATTT	−100	5·92	
togT	Oligogalacturonide transporter	tAATGAAACATTGTTTtAaTT	−24	5·90	
ogl	OGA lyase	AAATGAAACATcGTTTCtaTa	−98	5·80	
kdgM	OGA-specific porin	AAATGAAACATTGTTTCtTTg	−113	5·73	
ygjV	?	AtATGAAACggTGTTTtATTT	−149	5·70	
pemA	Pectin methylesterase A	AAATGgAAtATcGTTTCATTT	−115	5·59	
pelW-togMNAB	Oligogalacturonides transport and catabolism	gAtTGAAACgATGTTTCtaTT	−75	5·44	

Table 3. cont.

Operons	Function	Site	Position	Score	Regulation
*dhfX**	?	AAtTGAAACtATGTTTtgaTT	−77	5·34	
pehX	ExoPGA lyase	ttATaAAACAcTGTTTCtTTa	−38	5·24	
kdgK	KDG kinase	AttTaAAACgTcGTTTtAaTa	−92	5·23	
*kdgN-pelP-sghX**	OGA porin, pectate lyase, ?	AAATagAAtggcGTTTCATaa	−77	4·87	
*spiX**	? ~ Sugar isomerases	ttATaAAACATccTTcCATTa	−118	4·43	
Y. pestis					
togT	Oligogalacturonide transporter	AAATGAAACATTGTTTtAaTT	−24	6·02	
kduI-kduD/kdgF	DK-I catabolism	tAATaAAACATcaTTTCATTT	−200	5·92	
yjgK	?	tAATaAAACAgcaTTTCATTT	−40	5·84	
*kdgX**	? Predicted KDG permease	tAAcGAAACATcaTTTCATTT	−42	5·83	
ogl	OGA lyase	AAATGAAACATTGTTTCtaTa	−97	5·81	
pemA	Pectin methylesterase A	AAATGgAAtggcGTTTCATTT	−150	5·50	
kdgM	OGA-specific porin	AAATGAAACATTGcTTCtTTT	−129	5·48	
*kdgN-pelP-sghX**	OGA porin, pectate lyase, ?	AAATaAAAtAcTGTTTCATaa	−74	5·48	
pelW-togMNAB	Oligogalacturonides transport and catabolism	gAtTGAAACgATGTTTCtaTT	−75	5·44	
kdgK	KDG kinase	AttTaAAACAccGTTTtAaTc	−226	4·94	
yeeO	~ Multidrug efflux transporter	tcATGAAACtATGTTcCATTT	−159	4·74	
ygjV	?	AAAaaAtACgccaTTTCgTTa	−155	4·78	
		AtATGAAACggTGTTctgTTT	−134	4·88	
*spiX**	? ~ Sugar isomerases	AggTaAAACATatTTTCATTa	−121	4·51	
K. pneumoniae					
kduD	DK-I catabolism	AAATGAAACATcGTTTtAaaT	−208	5·64	
		AAAcGgAACtcTGTTTtATTT	−140	4·96	
ogl	OGA lyase	tAATGAAACAATGTTTtgaTa	−102	5·55	
kduD2	DK-I catabolism	AAtcaAAACATTGTTTCATTT	−105	5·79	
kdgF	?	AAATGAAACAATGTTTCAcTT	−162	5·77	
kdgM	OGA-specific porin	AAAcGAAAtgTcaTTTtATTT	−38	5·69	
		AAATaAAAtgATGTTTCATaa	−59	5·55	
yjgK	?	AAATGAAAtgcTGTTTtATaT	−30	5·59	
garD	D-Galactarate utilization	AAATGAAttAgcGTTTtATTT	−53	5·37	
togMNAB	Oligogalacturonides transport	AAtTaAAACggTGTTTtATaa	−209	5·30	
		tttTGAAACATcGTTTCgacT	−286	5·12	
yeeO	~ Multidrug efflux transporter	AttTGAAAtgTTGTTTCAaca	−8	5·24	
kdgK	KDG kinase	tAATGgAACgATGTTTtAaTa	140	5·18	
ppsA/ydiA	Phosphoenolpyruvate synthase/?	AAtTGAAACgcTaTTTttgTT	−253	5·06	
yjcB	? Small protein (94 a.a.)	AAcaGAAAtggcGTTTCATca	−119	5·04	
togT	Oligogalacturonides transport	tAtaaAAACAgTGTTTCAaaa	−45	4·96	
*gntDBMNAC**	?	tAtTGAAACgTatTTTCAaTa	−82	4·94	
pykF	Pyruvate kinase I, fructose-stimulated	AttTGAAACgAgGTTTttaTT	−283	4·79	
sotA	Sugar efflux transporter	tttTGAAACgcctTTTtAaTT	−103	4·77	
rhiT-rhiN	Rhamnogalacturonides catabolism	AAgaGAAAtgcTGTTTtATaa	−198	4·71	
*spiX**	? ~ Sugar isomerases	tttTaAAACAActTTTCATaa	−110	4·51	
S. typhimurium					
kduI-kduD	DK-I catabolism	AAATGAAACgTTGTTTtATTT	−85	6·15	
		AAtcaAAACAgTGTTTtgaTT	−20	5·20	
*kdgX** (*yifZ*)	? Predicted KDG permease	AAATaAAACATTaTTTtAaTT	−35	5·77	
yjgK	?	AAATaAAACgcTGTTTtAacT	−33	5·62	
yjcB	? Small protein (94 a.a.)	AAtaGAAACAATGTTTCATTc	−119	5·59	
kdgM	Rhamnogalacturonide-specific porin	AcATGAAACAAcGTTTtAaTT	−29	5·47	
rhiN	Rhamnogalacturonan catabolism	AAtaGAAACgcTGTTTtATaa	−188	5·21	
yeeO	~ Multidrug efflux transporter	tAtTGAAAtgTgGTTTCAacT	−6	5·13	
*spiX**	? ~ Sugar isomerase	tAATGAAACcTgGTTTtATTa	−107	4·96	
*tpfX**	?	ttATGAAACATTGTTTCAgaT	−230	4·91	

Table 3. cont.

Operons	Function	Site	Position	Score	Regulation
kdgK	KDG kinase	tAATGgAcCgATGTTTtAaTa	−62	4·59	
ppsA/ydiA	Phosphoenolpyruvate synthase/?	AAtgaAAAgtgTGTTTCATTT	−227	4·58	
***E. coli* K-12 and CFT073**					
kduI-kduD	DK-I catabolism	AAATGAAACATTGTTTtATTT	−129	6·16	
		AAAcGAAACAgTGTTTCAcTa	−62	5·34	
yjgK	?	AAATGAAACgTTGTTTtAaTT	−27	6·01	
kdgT	KDG transporter	AAATaAAACAgcGTTTCAaTT	−132	5·93	
ppsA/ydiA	Phosphoenolpyruvate synthase/?	tAtgaAAACAgcaTTTCATTT	−221	5·17	
kdgK	KDG kinase	tAATGgAACAcTGTTTtAaTa	132	5·11	
ygjV	?	AtAaaAAACggcGTTTCATaa	−63	4·97	
***E. coli* CFT073**					
ogl-kduD-kdgF-togT-pelX/spiX*	OGA lyase, DK-I catabolism, oligogalacturonides transport, pectate lyase/?	AttcaAAACATcaTTTCATTT	−140	5·30	
V. parahaemolyticus					
kduD-ygjV-kdgF-spiX-paeX/kdgR	DK-I catabolism, pectin acetylesterase/KDG repressor	TTTTaAAACGCCtTTTCAAAA	−122	6·06	
		TTTTaAAAtGCCtTTTCAAAA	−195	5·52	
*kdgM-pelX/togX1**	OGA porin, pectate lyase/? transporter	aaaTaAAAtGGtGTTTCAAAA	−21	5·21	
		aaTTaAAACaCCGTTTtAAAg	−111	5·67	
kdgK-kdgA	KDG kinase, KDG-6P aldolase	aaaTGAAcCGTtGTTTtAttt	−83	5·05	
togX2-ogl*	? Transporter, OGA lyase	aaTTaAAACaGCGTTTCgAgg	−30	4·87	
		aaaTGAAcCGTtGTTTtgAtt	−94	4·95	
V. vulnificus					
kduD1-ygjV1-kdgF-spiX/kdgR1	DK-I catabolism, pectin acetylesterase/KDG repressor	TTTTGAAACGTCtTTTCAAAA	−114	6·11	
		TTTTaAAAtaCCtTTTCAtAA	−188	5·16	
kdgK-kdgA1	KDG kinase, KDG-6P aldolase	aaTTGAAcCGTCGTTTtAAAA	−23	5·86	
kduD2-ygjV2-kdgK2-kdgA2/kdgR2	DK-I catabolism/KDG repressor (second copies)	TTTTaAAACGACGcTTtAAAA	−122	5·72	
*kdgM-pelX/togX1**	OGA porin, pectate lyase/transporter	aaacaAAACGGtGTTTCAAAA	−19	5·40	
		aaTTaAAACaGCGTTTCAAtg	−109	5·61	
VVA1633/VVA1634-35	Sulfatase activating protein/sulfatases	aaaTaAAAaGAtGTTTtAAAA	−66	5·45	
		TaTTGAAAtGAtGTTTCAAAA	−113	5·60	
*pelW-sghX**	ExoPGA lyase, ? ∼ glycosyl hydrolase	aTTaaAAACaTCGTTTCAAtt	−99	5·56	
VVA1631-30/VVA1632	?, ? ABC transporter (substrate-binding component)/sulfatase	aTaTaAAAaaGCGTTTCAAtA	−79	5·50	
		aaaTGAAAtGAtGgTTCAAtt	−185	4·81	
VVA1620-galTE-kduD3	? Transporter, galactose utilization, DK-I catabolism	aaaTGAAAaGCCGTTTtgttt	−356	4·81	
		TTTTGAAAtGAtGTTTtAtAA	−412	5·41	
ugl (*VVA1615*)	? ∼ Unsaturated glucuronyl hydrolase	aaTTGAAAaGGtGTTTtgAAA	−111	5·40	
VVA1628-1627-1626	? ABC transporter (transmembrane and ATPase components)	aacTGAAAaaGCGTTTtAAAA	−184	5·21	
togX2-ogl*	? Transporter, OGA lyase	aaTTaAAACaTCGTTTCgAgg	−27	4·87	
		aaaTGAAcCGTtGTTTtgAtt	−91	4·95	
cslA2 (*VVA1636*)	Chrondroitinase AC	aTaTaAAAttGCGTTTtAttt	−126	4·43	

and monomers (Ogl, KduI, KduD and KdgK). This set of genes agrees with the ability of these plant-pathogenic bacteria to degrade plant pectin and to use the resulting oligomers and monomers as a carbon source for growth. Among other bacteria considered in this study, only the *Yersinia* and *Vibrio* species possess pectinolytic enzymes, which are probably periplasmic since these bacteria lack orthologues of the Out system. These bacteria also contain cytoplasmic enzymes initially identified in *Erwinia* (Fig. 2). Thus, *Yersinia* and *Vibrio* are predicted to degrade pectic oligomers entering the periplasm via the KdgM porins, as well as unsaturated monomers. Considering the predicted

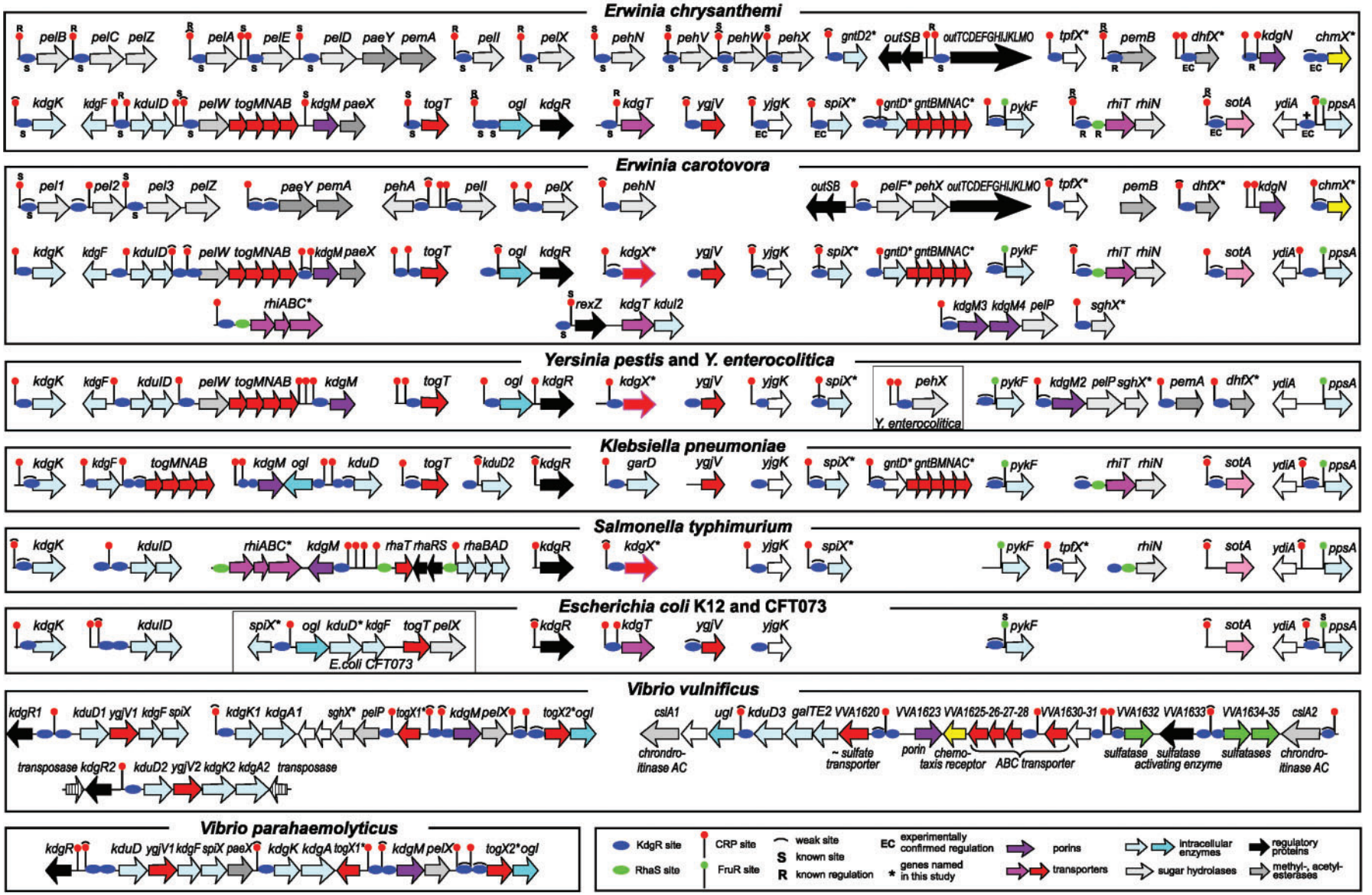

Fig. 2. Operon structures and predicted regulatory sites for the KdgR regulons in enterobacteria and *Vibrio* species.

composition of the KdgR regulon, *KP* should be able to degrade only short oligomers, dimers or trimers, by cytoplasmic enzymes (Fig. 2). *ST* and *EC* strain K-12, with only six predicted members of the KdgR regulon, appear to be able to degrade the unsaturated monomers DK-I, DK-II and KDG, but not pectic oligomers. Notably, *EC* strain CFT073 has acquired a cluster of genes allowing it to use short oligomers (Fig. 2). Thus, the *kduI*, *kduD* and *kdgK* genes involved in the degradation of unsaturated monomers constitute the core of the KdgR regulon conserved in all studied bacteria with the exception of *kduI*, which is absent in *KP* and the *Vibrio* species (possibly replaced by a non-orthologous enzyme, see below).

KdgR-regulated transporters

In *ER*, extracellular oligogalacturonides resulting from the pectin degradation first enter the periplasm by the specific porin KdgM (Blot *et al.*, 2002) and then cross the inner membrane using either an ABC transporter, TogMNAB or a GPH transporter, TogT (Hugouvieux-Cotte-Pattat & Reverchon, 2001). Genes required for the transport of oligogalacturonides (*kdgM*, *togMNAB* and *togT*) and their subsequent degradation to monomeric acid sugars in the cytoplasm (*pelW* and *ogl*) are present only in the *Erwinia*, *Yersinia*, *Klebsiella* and *Vibrio* species. *Erwinia* and *Yersinia* have from two to four homologues of the oligogalacturonate (OGA)-specific porin KdgM (Fig. 3). A close paralogue of *kdgM* in *ER* (*kdgN*) is preceded by a strong KdgR site. We have not detected candidate KdgR sites upstream of the *kdgM* gene in *ER*, although it was previously shown to be under KdgR regulation (Blot *et al.*, 2002). In addition to *kdgM* and *kdgN*, *EO* has two more homologues of OGA-specific porin located in one cluster with the periplasmic pectate lyase gene *pelP*. The upstream region of the possible *kdgM3-kdgM4-pelP* operon contains a candidate KdgR site. In both *Yersinia* species, there are two KdgR-regulated paralogues of the *kdgM* gene; the first one is located immediately after the *pelW-togMNAB* cluster, whereas the second one belongs to the possible *kdgM2-pelP-sghX* operon. Notably, the *sghX* gene (see *YPO3993* for reference), encoding a hypothetical secreted protein weakly similar to various glycosyl hydrolases, was also found as a single KdgR-regulated gene in *EO*. In addition, the *kdgK-kdgA* cluster in *VV* contains the *pelP-sghX* pair with a candidate KdgR site upstream. Orthologues of SghX were not found in other bacterial genomes. This genetic organization suggests that the function of SghX, a new member of the KdgR regulon that has a N-terminal signal sequence, may be closely related to the periplasmic pectate lyase PelP. KdgR-regulated *kdgM* homologues were also identified in *KP* and *ST*. However, *ST* lacks both known oligogalacturonide transport systems, TogT and TogMNAB. We propose rhamnogalacturonide specificity for the porin encoded by the KdgR-regulated *kdgM* homologue in *ST*, since this gene is located between genes possibly involved in rhamnogalacturonide transport (see below) and genes involved in the L-rhamnose catabolism (Fig. 2).

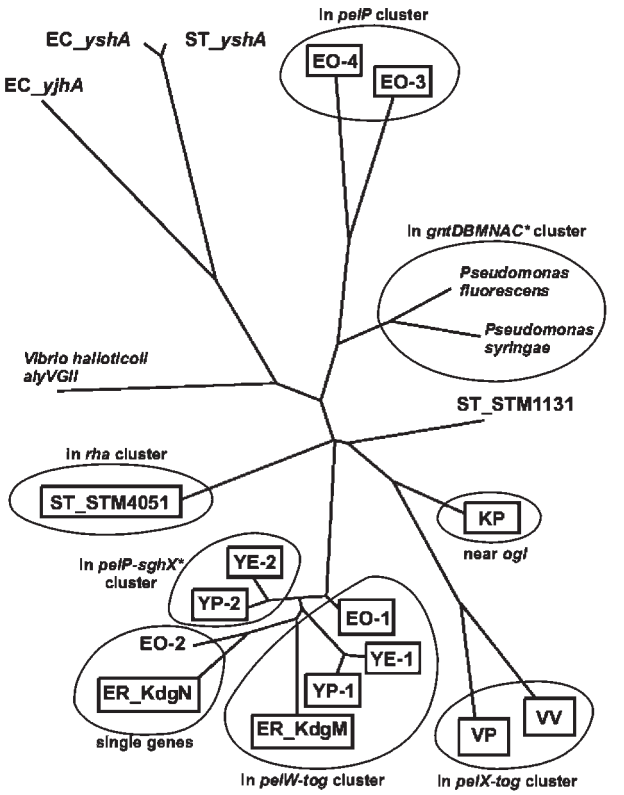

Fig. 3. Phylogenetic tree of bacterial orthologues of OGA-specific porin KdgM from *Erw. chrysanthemi*. Proteins are denoted by their genome abbreviations (listed in Methods). Multiple gene paralogues are numbered. Genes predicted to be regulated by KdgR are boxed.

Both TogMNAB and TogT, the oligogalacturonide transporters, are present in enterobacteria that are able to degrade pectic oligomers, namely *Erwinia*, *Yersinia* and *Klebsiella*. Only *togT* is present in the KdgR-regulated locus of the *EC* strain CFT073, allowing it to transport such oligomers. In contrast, the complete genomes of both *Vibrio* species, *VV* and *VP*, lack orthologues of both known OGA transport systems. In *Vibrio*, most KdgR-regulated genes are organized in one locus (Fig. 2) encoding porin KdgM, periplasmic pectate lyase PelX and cytoplasmic enzymes for the catabolism of OGAs (Ogl, KduD, KdgK, KdgA, etc.). This locus also contains two highly similar genes encoding hypothetical transporters with 14 candidate transmembrane segments (see *VVA1379* and *VVA1382*). The closest homologues of these two genes are sodium : glucose co-transporters from *Eukaryota*. To fill the metabolic gap, we tentatively assigned OGA specificity to these two transporters and named them TogX1 and TogX2.

Among enterobacteria, only *Erwinia* species and *EC* have the *kdgT* gene, which is a member of the KdgR regulon encoding a transporter for KDG, DK-I and DK-II (Condemine & Robert-Baudouy, 1987). Interestingly, the *kdgT* gene of *EO* is located in one putative operon with a

kduI paralogue (72 % similarity) and the *rexZ* gene encoding a regulator of exoenzyme production (Thomson *et al.*, 1999). In an attempt to identify the apparently missing KDG permease in other enterobacteria, we detected a candidate KdgR-regulated gene, named *kdgX* (previous *ST* name *yifZ*), which is present in *Yersinia* species, *ST* and *EO* (Fig. 2). KdgX has nine predicted transmembrane segments and belongs to the Drug/Metabolite transporter family. One characterized member of this family, RhaT of *EC*, functions in sugar uptake. Since all identified *kdgX* genes are preceded by candidate KdgR- and CRP-binding sites, we proposed that the specificity of the KdgX transporter is similar to that of KdgT in *ER* and *EC*.

The KdgR-regulated pectinases

Erwinia species possess a variety of extracellular pectinolytic enzymes, most of which are controlled by KdgR (Fig. 2). Both *ER* and *EO* contain single pectinase genes *pelI*, *pelX*, *pehN* and *pemB*. Two intracellular pectinase genes, *pelW* and *paeX*, are included in the large gene cluster encoding transporters and enzymes for OGA catabolism. There are several remarkable differences between the two *Erwinia* species concerning the arrangement of pectinolytic genes. The protein export system for secretion of extracellular pectinases is encoded by the *out* gene cluster and regulated by KdgR in *ER* (Condemine *et al.*, 1992). The *out* cluster in *EO* is also predicted to belong to the KdgR regulon and it includes the polygalacturonase gene *pehX* and a new pectate lyase gene, named *pelF*, which is most similar to the pectate lyase gene *pel* from *Bacillus subtilis* (Nasser *et al.*, 1993). While *EO* has only one *pehX* gene in the *out* cluster, the *pehX* gene of *ER* forms a cluster with two close paralogues, *pehV* and *pehW*, suggesting recent gene duplication in *ER*. The pectin acetylesterase gene *paeY* and the methylesterase gene *pemA* in *ER* are located in a KdgR-regulated operon with the pectate lyase gene *pelD*, whereas *EO* has a possible operon, *paeY-pemA*, with two upstream KdgR sites. It is noticeable that *EO* has no orthologue for *pelD* while the same *ER* cluster has two *pelD* paralogues, *pelA* and *pelE*, again suggesting gene duplication in *ER*. Finally, the *pelBCZ* gene cluster of *ER* contains two orthologous pectate lyase genes, *pelB* and *pelC*, preceding a weakly similar gene, *pelZ*. The same gene cluster of *EO* includes three close homologues of *pelB*/*C*, named *pel1*, *pel2* and *pel3*, and a *pelZ* orthologue. This organization suggests that an ancestor of the *pelB*/*C* genes was subject to one duplication event in *ER* and two duplications in *EO*. Duplication of pectinase genes seems to be a common phenomenon in *Erwinia* species and could favour adaptation of these pathogenic bacteria to various plant tissues.

Among other analysed bacteria, only *Yersinia* and *Vibrio* species have several pectinolytic enzymes (PelX, PehX, PelP, PemA, PaeX and possibly SghX), which also are members of the KdgR regulon. Although the Out-dependent secretion system is absent in these species, all these proteins contain candidate N-terminal signal sequences, arguing for their periplasmic location.

Other KdgR-regulated genes

A search for candidate KdgR-binding sites in bacterial genomes complemented by operon structure analysis allowed us to detect a number of new members of the KdgR regulon (Table 3). In *ER*, genes *ygjV*, *tpfX*, *chmX*, *ppsA*/*ydiA* (divergent genes), *gntD2*, *yeeO*, *spiX*, *yjgK* and *dhfX* are predicted to have strong KdgR-binding sites, i.e. sites with scores higher than 5·20. Most of these genes encode hypothetical proteins of unknown function. The *ppsA* product was previously characterized as a phosphoenolpyruvate synthase (Niersbach *et al.*, 1992), but its potential KdgR regulation has yet to be described. At this stage, we performed experimental verification of the predicted regulation for each novel candidate member of the KdgR regulon in *ER* prior to clarification of their role in pectin catabolism by detailed functional, positional and phylogenetic analysis of these genes.

We observed that the presence of a weak KdgR-binding site (score below 5·2, Table 3) can also have a biological significance. Indeed, some previously described genes, namely *pehX*, *pehW*, *pelI*, *rhiTN* and *pehV*, that are known to be controlled by KdgR, have KdgR sites with scores below cut-off (4·79, 4·64, 4·62, 4·37 and 4·35, respectively). Thus, we also tested potential KdgR regulation of genes that have a KdgR site conserved in other bacteria (*gntDBMNAC*, *pykF* and *sotA*) or whose function could be related to pectin catabolism or plant infection (*indA*, *pecT*, *pir* and *expI*).

Construction of *uidA* transcriptional fusions in novel members of the KdgR regulon

To analyse expression of the selected *ER* genes, namely *chmX*, *dhfX*, *gntB*, *gntD*, *gntD2*, *ppsA*, *pykF*, *sotA*, *spiX*, *tpfX*, *ydiA*, *yeeO*, *ygjV* and *yjgK*, we constructed transcriptional fusions by inserting a *uidA*-Km cassette into a selected restriction site located in the corresponding ORF. For the genes *indA*, *pecT*, *pir* and *expI*, we used previously constructed fusions (Table 1). Fourteen selected genes were cloned after PCR amplification (only their 5′ end for genes longer than 1 kb). The *uidA*-Km cassette was inserted into a restriction site situated in the coding region of each independently cloned gene (Table 1). Insertion of the cassette in the correct orientation generates a transcriptional fusion with the *uidA* gene encoding β-glucuronidase (Bardonnet & Blanco, 1992).

Plasmids containing the insertions were introduced into *ER* strain 3937. Transformants were submitted to successive cultures in Km-containing low-phosphate medium lacking the antibiotic to which resistance is encoded on the plasmid. Chromosomal *uidA*-Km insertions were obtained for the 11 genes *chmX*, *dhfX*, *gntB*, *ppsA*, *pykF*, *sotA*, *spiX*, *tpfX*, *ydiA*, *yeeO* and *yjgK*. In each case, the correct insertion of the cassette into the chromosome was confirmed by PCR. Attempts to obtain recombination of the insertions in three genes, *yjgV*, *gntD* and *gntD2*, were unsuccessful. Mutations

in these genes could be deleterious for the bacterial growth, preventing their isolation.

The growth of the 11 mutants was analysed using GA or PGA as sole carbon source. None of the mutants was affected for growth rate or the final growth yield of cultures with these compounds (data not shown). Thus, genes *chmX*, *dhfX*, *gntB*, *ppsA*, *pykF*, *sotA*, *spiX*, *tpfX*, *ydiA*, *yeeO* and *yjgK* are dispensable for PGA or GA catabolism. Since the *chmX* product could be involved in chemotaxis, we tested chemotactic ability towards GA and oligogalacturonides by determining the diameter of the rings observed on semi-solid agar plates. The size of the rings observed for the *chmX* mutant and parental strain A350 were similar when the medium was supplemented either with GA or oligogalacturonides (data not shown). This result suggests either that ChmX is not involved in chemotaxis towards GA and oligogalacturonides, or that *ER* possesses additional chemotaxis receptor proteins that are also specific to these compounds. Identification of two additional *chmX* homologues preceded by candidate KdgR sites (see below) reinforces the second hypothesis, arguing for the existence of at least three KdgR-regulated proteins that could be involved in chemotaxis towards pectic oligomers or monomers.

Expression of *uidA* transcriptional fusions in the candidate members of the KdgR regulon

We tested expression of the constructed fusions in the presence of various carbon sources and in different *ER* genetic backgrounds. The basal level of expression was determined in the presence of glycerol as carbon source. GA and PGA were used as potential inducing compounds. The fusions were transduced into *ER* strains A1077 and A576 that contain a mutation in *kdgR* and *kdgK*, respectively. The *kdgR* mutation allows for the direct determination of *in vivo* regulation of the fusion by KdgR. In the *kdgR* mutant A1077, expression of genes negatively controlled by KdgR, such as those encoding pectate lyases, increased in the absence or presence of the inducer. In the presence of PGA or GA, the *kdgK* mutation allows for accumulation of the KdgK substrate, KDG, which is the intracellular inducer interacting with KdgR. Accumulation of KDG in a *kdgK* mutant leads to a very high induction of genes controlled by the KdgR/KDG couple. For instance, the pectate lyase activity greatly increased in the *kdgK* mutant A576 in the presence of either PGA or GA. Thus, each fusion was assayed in the wild-type background, in the presence of a *kdgR* or *kdgK* mutation, in media supplemented with glycerol, GA or PGA. Based on the expression profiles of the tested genes, four classes of genes were defined (Fig. 4).

Seven genes of class 1, namely *chmX*, *dhfX*, *gntB*, *pykF*, *tpfX*, *yeeO* and *yjgK*, were moderately induced by GA or PGA in the wild-type background, with induction ratios of two- to sixfold. The induction ratios of *chmX*, *dhfX* and *ygjK* clearly increased in the *kdgK* mutant, reaching 10- to 30-fold. Moreover, they showed a highly derepressed expression in the *kdgR* mutant, with a β-glucuronidase activity 16- to 26-fold higher than that of the wild-type in the uninduced medium (Fig. 4). The expression of *gntB*, *pykF*, *tpfX* and *yeeO* remained moderately induced in a *kdgK* mutant and was derepressed in the *kdgR* mutant by factors of 3, 20, 6 and 5, respectively. Thus, KdgR clearly represses expression of *chmX*, *dhfX*, *gntB*, *pykF*, *tpfX*, *yeeO* and *yjgK*. Most previously described genes of the KdgR regulon showed similar results (data not shown) and, on the basis of their expression, the following genes could be considered as members of class 1: *kdgT*, *kduI-kduD*, *togT*, *pelW-togMNAB* and *pelX*. It was noticed that all genes of class 1, except *pykF* and *yeeO*, have strong KdgR sites with scores between 6 and 5·2 (Table 3). However, some genes with high scoring, mainly extracellular pectate lyase genes *pelE*, *pelA*, *pelC-pelZ*, *pelB* and *pelD* (scores between 5·6 and 5·2), do not belong to this class. It is known that expression of the *pel* genes is controlled by a set of regulators (KdgR, PecS, PecT, CRP, etc.) and the direct effect of KdgR is probably modulated by competition between the regulatory proteins for binding to adjacent sites.

Two genes constituting class 2, *spiX* and *sotA*, were not significantly induced by GA or PGA in the wild-type background. However, their transcription was stimulated three- to fivefold in the *kdgK* mutant in the presence of GA or PGA. Transcription of these two genes was also derepressed in the *kdgR* mutant; the fusion expression increased about threefold compared to the wild-type strain under non-inducing conditions. These data indicate that *spiX* and *sotA* are weakly controlled by the KdgR repressor. The scores of the predicted KdgR sites of *spiX* and *sotA* are 5·3 and 4·6, respectively. Previously characterized members of class 2, i.e. genes moderately controlled by KdgR, include *pehX*, *pehW*, *pehV* and *rhiT-rhiN* (scores between 4·8 and 4·3). These genes are known to be only partially regulated by KdgR and other regulators are involved in their expression; for instance, RhaS is the main activator of *rhiT-rhiN* transcription (Hugouvieux-Cotte-Pattat, 2004). Additional unidentified regulators could have a major role in the control of *spiX* and *sotA* transcription.

Two divergently transcribed genes, *ppsA* and *ydiA*, share a common regulatory region containing a strong predicted KdgR-binding site (score 5·6, Table 3). We observed an entirely different expression profile for these genes (Fig. 4) and assigned them to class 3. Both genes were weakly repressed in the presence of GA or PGA in the wild-type background, with repression ratios of about two. This repression became more visible in the *kdgK* mutant, reaching three- to sixfold. Moreover, both genes showed decreased expression in the *kdgR* mutant. These results are typical for positive regulation rather than negative control. Thus, KdgR behaves as an activator of *ppsA* and *ydiA* expression. Intracellular formation of KDG during pectinolysis provokes dissociation of KdgR from its operators and in such conditions expression of *ppsA* and *ydiA* decreases.

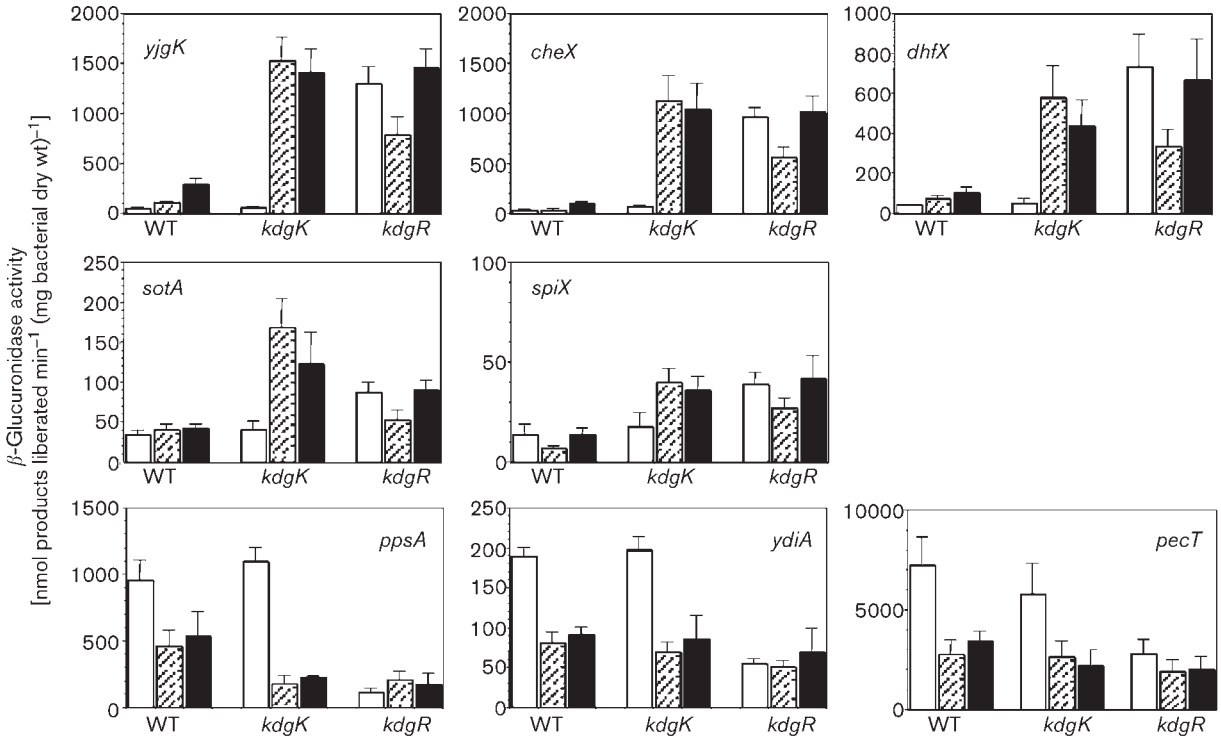

Fig. 4. Expression of transcriptional fusions in the predicted members of the KdgR regulon. Strains containing a *uidA* transcriptional fusion in one of the selected genes were grown in minimal medium containing either glycerol (no inducer, white bars) or GA (hatched bars) or PGA (black bars) as a potential inducer. Each fusion was introduced in a wild-type background (WT) and in mutants affected either for the gene *kdgK*, accumulating the intracellular inducer KDG, or for the regulatory gene *kdgR*. The β-glucuronidase activities reported are the means of three to five independent experiments and standard deviations are indicated.

There are several examples of regulators that could be either activator or repressor (the most classical one being CRP), but this double function was not previously observed in the case of KdgR.

We analysed the role of KdgR in the expression of transcriptional fusions in *indA*, *pecT*, *pir* and *expI* that have candidate KdgR sites with scores of 4·95, 4·84, 4·76 and 4·65, respectively. The fusions were transduced into either *kdgR* or *kdgK* mutant strains and their expression was measured under non-inducing or inducing conditions. The *pecT* gene appeared to be weakly positively regulated by KdgR and thus could be considered as a member of class 3 (Fig. 4). In contrast, the great variability of expression of *indA* or *expI* did not allow us to observe a significant reproducible effect of a *kdgR* mutation on their expression (data not shown). Expression of the *pir* gene was clearly independent of KdgR (data not shown). In this case, the site detected as a potential KdgR-binding site could be a Pir-binding site. Indeed, Pir is also a regulator of the IclR family and its binding site was shown to be similar to that of KdgR (Nomura *et al.*, 1999). This observation prompted us to verify that the expression of the genes *chmX*, *dhfX*, *expI*, *indA*, *pecT*, *ppsA*, *sotA*, *spiX*, *ydiA* and *yjgK* is not affected by Pir (data not shown).

Potential function of genes strongly regulated by KdgR: *chmX*, *dhfX* and *yjgK*

Enterobacteria possess a set of methyl-accepting chemotaxis proteins (MCPs) which are involved in the control of flagellar activity so that the bacterial cells move toward favourable environmental conditions (Stock & Surette, 1996). The periplasmic substrate-binding component TogB of the TogMNAB transport system acts as an oligogalacturonide-specific chemoreceptor in *ER* (Hugouvieux-Cotte-Pattat *et al.*, 2001). Thus, TogB most probably interacts with an oligogalacturonide-specific MCP of the inner membrane, which could transduce the signal to the motility apparatus. A newly identified KdgR-regulated MCP gene, *chmX*, is a good candidate for this function. The absence of an observable phenotype of the *chmX* mutant suggests that additional MCP proteins allow *ER* to be attracted by pectic oligomers. A large number of MCP homologues observed in the *ER* chromosome suggests possible redundancy in their substrate specificity. Among 44 predicted MCP genes, two more genes (*chmX13* and *chmX21*) are preceded by candidate KdgR-binding sites with scores only slightly below the cut-off (5·09 and 4·97, respectively; Table 3). Thus, additional experiments are necessary to conclusively assign the role of ChmX, ChmX13 and ChmX21 in chemotaxis towards pectic oligomers or monomers.

In both the *Erwinia* and *Yersinia* species, the KdgR regulon includes a hypothetical protein, DhfX, from the dienelactone hydrolase family (see GenBank entry NP_667845 for reference). DhfX has no other orthologues and is weakly similar to an acetyl xylan esterase from *Bacillus pumilus* and a cephalosporin C deacetylase from *B. subtilis*. Since KdgR-regulated protein DhfX of *ER* has a candidate N-terminal signal sequence, we propose that it is a periplasmic esterase acting on pectic oligomers, possibly a novel pectin acetyl esterase, in addition to PaeY and PaeX.

Hypothetical gene *yjgK* was predicted to be regulated by KdgR in all enterobacteria. In *ER*, it is among the genes which are strongly controlled by KdgR *in vivo*. We noticed that the *yjgK* gene from *YE*, in addition to the predicted KdgR site, has a candidate binding site for the ExuR repressor, a regulator of the GA catabolism. The *yjgK* product belongs to the DUF386 family, consisting of conserved hypothetical proteins, typically about 150 aa in length, with no known function. The phylogenetic tree of this family has several distinct branches, three of which, YjgK, YhcH and YiaL, are specific for enterobacteria (data not shown). While *yjgK* is a single gene in all enterobacteria, two other *EC* members of the DUF386 family are located in gene clusters involved in the catabolism of *N*-acetylneuraminic acid (*nanATKE-yhcH*) and, possibly, 2,3-diketo-L-gulonate (*yiaKLMNOPQRS*) (Yew & Gerlt, 2002). A more sensitive homology search with PSI-BLAST showed weak similarity of YjgK to EbgC of *EC* (17 % identity, 21 % similarity). The function of EbgC is not well defined, but it is required for the full activity of the second *EC* β-galactosidase encoded by the *ebgA* gene (Elliott *et al.*, 1992). Since YjgK is one of the most conserved members of the KdgR regulon, we suppose that it may be involved in the downstream part of the pectin catabolic pathway, probably being required for full activity of a conserved enzyme, KduD, KduI or KdgK.

The KdgR-regulated genes *tpfX* and *yeeO*

In contrast to all other KdgR-regulated genes, *tpfX* and *yeeO* lack candidate binding sites for the catabolic repressor protein CRP. Orthologues of the *tpfX* gene were found only in two other enterobacteria, *EO* and *ST*, where they are also predicted members of the KdgR regulon (see *STM1931* in *ST*). Hypothetical protein TpfX belongs to the ThiJ/PfpI family that includes thiamine biosynthesis protein ThiJ from bacteria and intracellular protease PfpI from archaea. Although orthologues of *yeeO* were found in all enterobacteria, predicted KdgR-binding sites were observed only in *Erwinia* species, *KP* and *ST*. This gene encodes a hypothetical transport protein from the multi-antimicrobial extrusion family (Hvorup *et al.*, 2003). The data available are insufficient to assign a role to *tpfX* and *yeeO*.

Genes weakly regulated by KdgR: *spiX* and *sotA*

In all enterobacteria, the predicted KdgR regulon includes a hypothetical sugar isomerase gene *spiX* (see *STM1933* in *ST* for reference). In *ER*, this gene is expressed at a low level and is weakly controlled by KdgR (Fig. 4). Searching the databases, we identified SpiX orthologues in other bacterial species (the phylogenetic tree is shown in Fig. 5). In the *Vibrio* species, the *spiX* gene is located in the KdgR-regulated cluster *kduD-ygjV-kdgF-spiX* (Fig. 2). Notably, the complete genomes of *VV* and *VP*, as well as the unfinished genome of *KP*, lack the *kduI* gene involved in the first step of pectic monomer catabolism, isomerization of DK-I to DK-II. In some bacteria from the *Bacillus*/*Clostridium* group, the *spiX* orthologues are located in the cluster, including *kdgK*, *kdgA* and *kduD* (Fig. 5), and there are no *kduI* homologues in these genomes. However, in *Xanthomonas* species, we observed the *kduI-kduD-spiX* gene cluster encoding both the KduI and SpiX isomerases. In contrast to *EC* strain K-12, *EC* strain CFT073 possesses an additional KdgR-regulated locus, including the *spiX* gene (Fig. 2). A search at a low level of stringency allowed us to observe distant homology of SpiX to galactose-6-phosphate isomerase LacB from *Streptococcus mutans* and ribose-5-phosphate isomerase RpiB from *EC*. Summarizing all these data for SpiX, we propose that this novel member of the KdgR

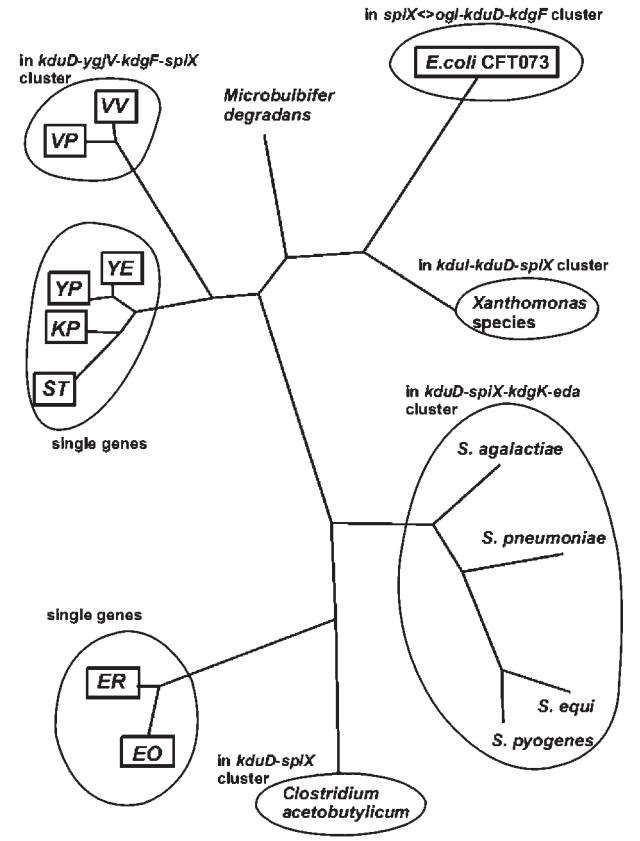

Fig. 5. Phylogenetic tree of hypothetical sugar isomerases, SpiX. Proteins are denoted by their genome abbreviations (listed in Methods). Genes predicted to be regulated by KdgR are boxed.

regulon could function as an additional isomerase, complementing the absence of KduI in some bacterial species.

Another gene weakly regulated by KdgR encodes the sugar efflux transporter SotA (Condemine, 2000). Expression of *sotA* in *ER* is activated by CRP and weakly induced by GA. The *sotA* gene is preceded by a weak KdgR site in both the *Erwinia* and *Klebsiella* species, whereas the *Yersinia* species lack the *sotA* gene. In *ST* and *EC*, this gene is present but has no candidate KdgR site in upstream region (Fig. 2). Conservation of a candidate CRP site upstream of *sotA* in all enterobacteria is in agreement with the previously proposed broad substrate specificity of this sugar efflux pump. Weak regulation of SotA by KdgR confirmed in *ER* (Fig. 4) indicates that in plant-pathogenic bacteria, SotA could be more specifically involved in the efflux of intermediates of pectin catabolism that could have a toxic effect if they accumulated intracellularly. Indeed, a strong toxic effect was observed in *EC* for 6-phospho-KDG (Fuhrman *et al.*, 1998) and growth inhibition was frequently observed in *ER* mutants accumulating DK-I, DK-II or KDG (unpublished observations).

Function of genes positively regulated by KdgR: *ppsA*, *ydiA* and *pecT*

In both *Erwinia* species, a strong KdgR site was identified in the common upstream region of two divergently transcribed genes, *ydiA* and *ppsA*, encoding a hypothetical conserved protein of unknown function and phosphoenolpyruvate synthase, respectively. Although the *ydiA*/*ppsA* gene cluster was identified in all enterobacteria, the predicted KdgR-binding site is not conserved in *Yersinia* or *Salmonella* species (Fig. 2). Since expression of both these genes is reduced in *ER* in the presence of GA or PGA, and in the *kdgR* mutant, we concluded that they are positively regulated by KdgR. All previously known members of the KdgR regulon are negatively regulated by this transcriptional factor (Hugouvieux-Cotte-Pattat *et al.*, 1996). In *EC*, the *ppsA* gene is positively regulated by FruR, a global regulator of the carbon utilization (Negre *et al.*, 1998). Using the FruR site profile, we showed that the candidate FruR-binding site in the *ydiA*/*ppsA* regulatory region is conserved in all enterobacteria (Fig. 2). In most bacteria, this region also contains a CRP-binding site (Fig. 2). In *EC*, expression of *ppsA* is also negatively regulated by the carbon storage regulator CsrA (Sabnis *et al.*, 1995). Thus, complex regulation of the *ppsA* gene could take place in *Erwinia* species, involving several regulators of sugar catabolism: KdgR, FruR, CRP and possibly CsrA (RsmA in *EO*). The catabolically activated phosphoenolpyruvate synthase PpsA is a key gluconeogenic enzyme in *EC* (Oh *et al.*, 2002). The metabolic role of the *ydiA* gene product is not clear; it could also be linked to gluconeogenesis since *ydiA* is co-localized and probably co-regulated with *ppsA* in all enterobacteria. We conclude that the role of KdgR, at least in *Erwinia*, is not restricted to the negative control of the pectin catabolism, but is extended to the positive regulation of gluconeogenesis. The effect of KdgR will be to favour carbon flow through the gluconeogenic pathway when pectin is not metabolized. Thus, during plant infection, KdgR could play a role in coordination of central carbohydrate metabolism by directing the intracellular carbon flux. This role could be even larger since we noticed that among weaker candidate KdgR sites (score 4·75), there is the *pykF* gene encoding fructose-stimulated pyruvate kinase I. Regulation of *pykF* was shown to be opposite to that of *ppsA*, since this gene is involved in glycolysis. In *EC*, *pykF* is repressed by FruR and activated by CsrA (Bledig *et al.*, 1996; Sabnis *et al.*, 1995). The *pykF* upstream regions in *ER*, *EO*, *KP* and *EC* contain candidate KdgR sites with scores ranging from 4·68 to 4·95. Exactly the same set of enterobacterial genomes is predicted to have a KdgR-regulated *ppsA* gene. The respective position of candidate binding sites and promoter elements in the *pykF* upstream regions of these enterobacteria suggests negative regulation of *pykF* by both KdgR and FruR. Our *in vivo* analysis confirmed that PpsA and PykF, catalysing reverse reactions of the central carbohydrate metabolism, are regulated by KdgR in the opposite manner.

PecT is a negative regulator of the LysR family involved in the control of the pectate lyase synthesis (Surgey *et al.*, 1996). PecT expression is subject to autoregulation and negatively controlled by the nucleoid-associated protein H-NS (Nasser & Reverchon, 2002). The signal to which PecT responds remains unknown, but it is clear that variations in PecT concentration have drastic effects on the controlled genes. We showed that KdgR contributes to modulation of the PecT intracellular concentration, although the KdgR effect is weaker than that observed previously with the two negative regulators of *pecT* transcription, PecT and H-NS. The positive regulation exerted by KdgR could be an antirepressor effect rather than a direct activation. The regulatory network involved in the control of the pectate lyase synthesis includes several cross-relations. Identification of an additional link between KdgR and PecT adds a novel complexity between two pathways of this interactive network.

Potential function of other candidates of the KdgR regulon: *gntD*, *gntBMNAC* and *ygjV*

In both *Erwinia* species and *KP*, we found a new KdgR-regulated locus, named *gntDBMNAC*. The short distances between these genes make it likely that they form an operon. In addition, *ER* has a close paralogue of *gntD*, a single gene, *gntD2*, which is also preceded by a strong candidate KdgR-binding site (Fig. 2). While we could not obtain data for *gntD* and *gntD2*, we showed that *gntB* is controlled *in vivo* by KdgR. The *gntD* product is similar to two sugar acid dehydratases from *EC* that are specific to D-glucarate (GudH) and D-galactonate (DgoA). The *gntBMNAC* genes encode components of an ABC transport system from the oligopeptide permease family, including one substrate-binding protein, two transmembrane proteins and two ATP-binding proteins. The characterized members of this large family transport a variety of substrates, including small peptides, opines, nickel, α-galactosides

and other oligosaccharides (Gage & Long, 1998). A similar *gntDBMNAC* locus was found in *Pseudomonas syringae* pv. *tomato* and *Pseudomonas fluorescens*. Noteworthy, in contrast to *Pseudomonas aeruginosa* and *Pseudomonas putida*, these two plant-associated *Pseudomonas* species have the KDG kinase gene *kdgK* and a homologue of the OGA-specific porin *kdgM* (Fig. 3), although other pectin catabolic genes from *ER*, including *kduI* and *kduD*, were not found in these complete genomes. Moreover, *kdgM* and *gntDBMNAC* are divergently transcribed in *P. fluorescens*, whereas these genes probably form a single transcriptional unit in *P. syringae* pv. *tomato*. Considering these data, we propose that the function of the KdgR-regulated locus *gntDBMNAC* is the catabolism (GtnD) and active transport (GtnBMNAC) of some direct KDG precursor, most probably of plant origin.

The *ygjV* gene encoding a hypothetical transporter with four predicted transmembrane segments was found immediately downstream of the GA catabolic cluster *uxaCBA* in all enterobacteria except *ST*. Candidate KdgR sites upstream of this gene were observed in all enterobacteria, except *KP*. A KdgR-binding site is located between *uxaA* and *ygjV*, immediately after the predicted ρ-independent transcriptional terminator of the *uxa* operon. Moreover, we identified several paralogues of *ygjV* in both *Vibrio* species possessing the KdgR regulon, three paralogues in *VV* and two paralogues in *VP*. In both *Vibrio* species, one paralogue is located within the *uxaBC-kdgKA-ygjV* cluster (not regulated by KdgR), whereas another copy belongs to the *kduD-ygjV-kdgF-spiX* cluster preceded by two candidate KdgR sites (Fig. 2). The *Vibrio* species also have additional copies of the *kdgK* and *kdgA* genes that are members of the KdgR regulon. The duplication of these catabolic genes in *Vibrio* could be explained by a recent specialization of the paralogues towards catabolism of pectin (regulated by KdgR) or GA, two catabolic pathways converging on KDG formation. YgjV has no orthologues in other genomes and is not similar to other proteins from public databases. The predicted regulation by KdgR and clustering with the *uxa* genes suggest that genes of the *yjgV* family could be involved in transport of some intermediates of DK-I and GA catabolic pathways.

Conclusions

In this study we combined bioinformatic and experimental approaches to reconstruct and compare the pectin degradation pathways and the KdgR regulons in various gamma-proteobacteria. Fig. 6 summarizes previously known and newly identified members of the KdgR regulon and shows the main differences between the KdgR-regulated pathways in related gamma-proteobacteria. Two animal-associated bacteria, *EC* strain K-12 and *ST*, possess only the core part of this catabolic pathway, allowing them to utilize only monomers DK-I, DK-II and KDG. However, a recently sequenced uropathogenic strain of *EC* (CFT073) acquired an additional KdgR-regulated locus for transport and catabolism of short oligogalacturonides. *KP* also seems to use only short oligogalacturonides as KDG precursors. In contrast, *Yersinia* species possess two periplasmic pectinases, a pectate lyase and a polygalacturonase, and thus could utilize longer oligogalacturonide molecules. The KdgR regulons of two plant-pathogenic *Erwinia* species are the largest ones and contain an array of genes for the extracellular degradation of polymeric plant pectin and subsequent utilization of the resulting pectin oligomers of various lengths (Fig. 6).

Bacteria from another family, *Vibrionaceae*, also have KdgR regulons, although with several differences (Fig. 2). Both *Vibrio* species, *VV* and *VP*, contain a large cluster of KdgR-regulated genes encoding a porin and different enzymes necessary for the catabolism of OGAs. This locus also contains hypothetical transporters (TogX, Fig. 6) that could be responsible for OGA uptake in these species lacking TogT or TogMNAB homologues. Interestingly, *VV* has an additional KdgR-regulated locus encoding two homologues of chrondroitinase AC, a potential disaccharide ABC transporter, porin, three hypothetical sulfatases, sulfatase-activating enzyme, a homologue of a sulfate transporter and a homologue of the unsaturated glucuronyl hydrolase Ugl from *Bacillus* sp. We propose that this locus could be involved in the catabolism of chondroitin sulfate, a sulfated polysaccharide consisting of 1,4-linked derivatives of hexosamine and D-glucuronate. The predicted regulation of these genes by KdgR could be explained by the fact that the action of unsaturated glucuronyl hydrolase on chondroitin disaccharide will produce Δ4,5-D-glucuronate, which is spontaneously transformed into DK-I, a KdgR inducer (Hashimoto *et al.*, 1999).

This unexpected observation provides one more example of extension of the KdgR regulon in some bacterial species. For example, the KdgR regulon of *Erwinia* species is significantly extended to include most of the known pectin degradation enzymes, as well as the Out system for pectinase secretion. In addition, the KdgR regulon in *ER* includes the rhamnose-regulated operon *rhiTN* for transport and catabolism of rhamnogalacturonides (Hugouvieux-Cotte-Pattat, 2004). Double regulation of *rhiTN* by RhaS, activator of the rhamnose catabolism, and KdgR is explained by formation of both rhamnose and DK-I by cleavage of this oligosaccharide. In other enterobacteria (*EO*, *KP* and *ST*) possessing *rhiTN* or only *rhiN*, we observed conservation of RhaS- and KdgR-binding sites in their promoter region (Fig. 2). Moreover, a search for other RhaS and KdgR-regulated genes allowed us to identify a new TRAP-type transport system, named *rhiABC*, in *EO* and *ST* (Fig. 2). In *EO*, the *rhiABC* locus is preceded by candidate RhaS and KdgR sites. In the *ST* chromosome *rhiABC* has only a RhaS site, but it is adjacent to a KdgR-regulated gene, *kdgM*, itself adjacent to the rhamnose utilization locus *rhaT-rhaBAD*. RhiABC is a good candidate for the function of rhamnogalacturonide transporter, mainly in *ST*, in which RhiT is missing. Another example of possible

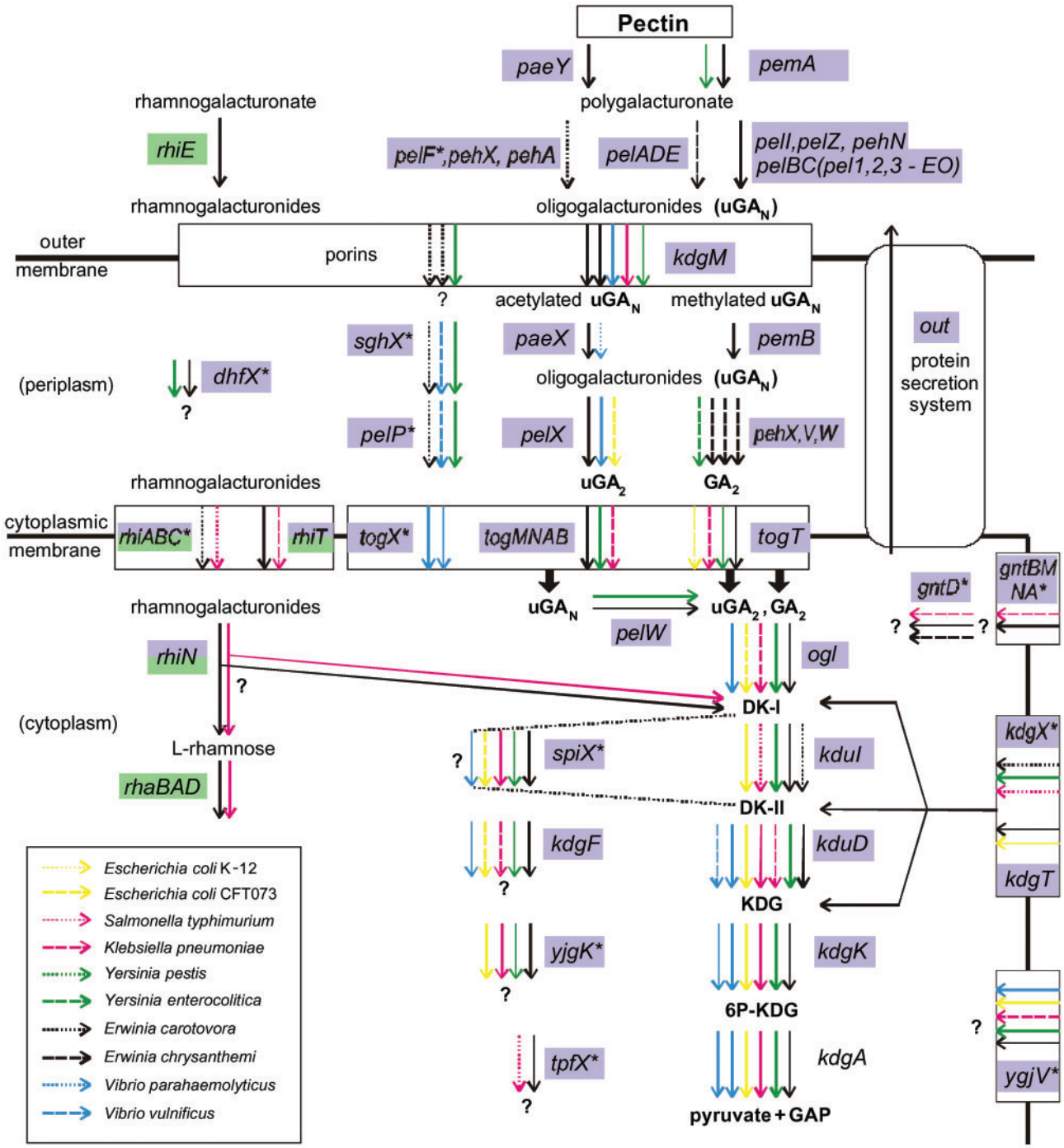

Fig. 6. Reconstruction of the catabolic pathway for pectin and its derivatives in gamma-proteobacteria. The KdgR- and RhaS-regulated genes are highlighted in blue and green, respectively. Solid arrows denote the presence of enzyme/transporter in both species according to the colour scheme shown. Newly identified KdgR-regulated genes are marked by asterisks.

regulon extension was observed in *KP*, which has a strong candidate KdgR site in the regulatory region of the divergently transcribed *garD*/*garPLRK* operons involved in D-galactarate catabolism. However, the significance of KdgR regulation of this pathway is not clear.

The use of comparative analysis allowed us to extend the knowledge about the KdgR regulon in the plant-pathogenic bacterium *Erw. chrysanthemi*, resulting in identification of ten novel genes preceded by strong KdgR sites (*chmX*, *dhfX*, *gntD*, *ppsA*, *spiX*, *tpfX*, *ydiA*, *yeeO*, *ygjV* and *yjgK*).

Experiments conducted to verify these predictions indicated that seven novel genes, *chmX*, *dhfX*, *gntB*, *spiX*, *tpfX*, *yeeO* and *yjgK* are indeed negatively controlled by KdgR. Predictions of sites with weaker scores also led us to analyse the expression of some previously identified genes, and we observed that regulator PecT, glycolytic enzyme PykF and sugar efflux transporter SotA are also regulated by KdgR, albeit at a lower level. Demonstration of the KdgR influence on the PecT intracellular level provides a new example of interactions between different regulators in the regulatory network controlling pectate lyase synthesis in *Erw. chrysanthemi*. Moreover, *pecT* and two genes with strong KdgR sites, *ppsA* and *ydiA*, were found to be positively regulated by KdgR. Noteworthy is the fact that the effect of gene activation has not been described previously for the classical repressor KdgR. Thus, the results presented here clearly demonstrate the interest of comparative genomics for the prediction of gene regulation, reconstruction of metabolic pathways and identification of apparently missing steps, either for transport systems or for enzymic activities.

ACKNOWLEDGEMENTS

The authors are grateful to members of the Lyon *Erwinia* group, Guy, Sylvie, Vladimir and William, for helpful discussions. We thank A. A. Mironov for providing software for genome analysis and useful discussions, and to O. Laikova for the RhaS recognition profile. This study was partially supported by grants from the Howard Hughes Medical Institute (55000309), Russian Foundation for Basic Research (02-04-49111), the Centre National de la Recherche Scientifique, the Ministère de l'Education Nationale et de la Recherche and the Programme Microbiologie 2003 (ACIM-2-17). This study has been done in part during the visit by D. R. to the Unit of Microbiology and Genetics, INSA-Lyon, France, supported by an exchange grant within the ESF Programme on Integrated Approaches for Functional Genomics.

REFERENCES

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Bardonnet, N. & Blanco, C. (1992). *uidA* antibiotic resistance cassettes for insertion mutagenesis, gene fusion and genetic constructions. *FEMS Microbiol Lett* **93**, 243–248.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000). GenBank. *Nucleic Acids Res* **28**, 15–18.

Bledig, S. A., Ramseier, T. M. & Saier, M. H., Jr (1996). FruR mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*. *J Bacteriol* **178**, 280–283.

Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A. & Condemine, G. (2002). The oligogalacturonate-specific porin KdgM of *Erwinia chrysanthemi* belongs to a new porin family. *J Biol Chem* **277**, 7936–7944.

Condemine, G. (2000). Characterization of SotA and SotB, two *Erwinia chrysanthemi* proteins which modify isopropyl-beta-D-thiogalactopyranoside and lactose induction of the *Escherichia coli lac* promoter. *J Bacteriol* **182**, 1340–1345.

Condemine, G. & Robert-Baudouy, J. (1987). 2-Keto-3-deoxygluconate transport system in *Erwinia chrysanthemi*. *J Bacteriol* **169**, 1972–1978.

Condemine, G., Dorel, C., Hugouvieux-Cotte-Pattat, N. & Robert-Baudouy, J. (1992). Some of the *out* genes involved in the secretion of pectate lyases in *Erwinia chrysanthemi* are regulated by KdgR. *Mol Microbiol* **6**, 3199–3211.

Elliott, A. C., Sinnott, M. L., Smith, P. J., Bommuswamy, J., Guo, Z., Hall, B. G. & Zhang, Y. (1992). The catalytic consequences of experimental evolution. Studies on the subunit structure of the second (ebg) beta-galactosidase of *Escherichia coli*, and on catalysis by *ebgAB*, an experimental evolvant containing two amino acid substitutions. *Biochem J* **282**, 155–164.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368–376.

Fuhrman, L. K., Wanken, A., Nickerson, K. W. & Conway, T. (1998). Rapid accumulation of intracellular 2-keto-3-deoxy-6-phosphogluconate in an Entner–Doudoroff aldolase mutant results in bacteriostasis. *FEMS Microbiol Lett* **159**, 261–266.

Gage, D. J. & Long, S. R. (1998). Alpha-galactoside uptake in *Rhizobium meliloti*: isolation and characterization of *agpA*, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. *J Bacteriol* **180**, 5739–5748.

Gelfand, M. S. (2003). Computational identification of regulatory sites in DNA sequences. In *Artificial Intelligence and Heuristic Methods in Bioinformatics*, pp. 148–172. Edited by P. Frasconi & R. Shamir. Amsterdam: IOS Press.

Gelfand, M. S. & Laikova, O. N. (2003). Prolegomena to the evolution of transcriptional regulation in bacterial genomes. In *Frontiers in Computational Genomics*, pp. 195–216. Edited by E. V. Koonin & M. Y. Galperin. Wymondham: Caister Academic Press.

Gelfand, M. S., Novichkov, P. S., Novichkova, E. S. & Mironov, A. A. (2000). Comparative analysis of regulatory patterns in bacterial genomes. *Brief Bioinform* **1**, 357–371.

Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S. & Murata, K. (1999). Unsaturated glucuronyl hydrolase of *Bacillus* sp. GL1: novel enzyme prerequisite for metabolism of unsaturated oligosaccharides produced by polysaccharide lyases. *Arch Biochem Biophys* **368**, 367–374.

Hofmann, K. & Stoffel, W. (1993). TMbase – A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler* **374**, 166.

Hugouvieux-Cotte-Pattat, N. (2004). The RhaS activator controls the *Erwinia chrysanthemi* 3937 genes *rhiN*, *rhiT* and *rhiE* involved in rhamnogalacturonan catabolism. *Mol Microbiol* **51**, 1361–1374.

Hugouvieux-Cotte-Pattat, N. & Reverchon, S. (2001). Two transporters TogT and TogMNAB are responsible for oligogalacturonide uptake in *Erwinia chrysanthemi* 3937. *Mol Microbiol* **41**, 1125–1132.

Hugouvieux-Cotte-Pattat, N. & Robert-Baudouy, J. (1989). Isolation of *Erwinia chrysanthemi* mutants altered in pectinolytic enzyme production. *Mol Microbiol* **3**, 1587–1597.

Hugouvieux-Cotte-Pattat, N., Dominguez, H. & Robert-Baudouy, J. (1992). Environmental conditions affect the transcription of the pectinase genes of *Erwinia chrysanthemi* 3937. *J Bacteriol* **174**, 7807–7818.

Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W. & Reverchon, S. (1996). Regulation of pectinolysis in *Erwinia chrysanthemi*. *Annu Rev Microbiol* **50**, 213–257.

Hugouvieux-Cotte-Pattat, N., Blot, N. & Reverchon, S. (2001). Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in *Erwinia chrysanthemi* 3937. *Mol Microbiol* **41**, 1113–1123.

Hvorup, R. N., Winnen, B., Chang, A. B., Jiang, Y., Zhou, X. F. & Saier, M. H. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *Eur J Biochem* **270**, 799–813.

James, V. & Hugouvieux-Cotte-Pattat, N. (1996). Regulatory systems modulating the transcription of the pectinase genes of *Erwinia chrysanthemi* are conserved in *Escherichia coli*. *Microbiology* **142**, 2613–2619.

Liu, Y., Jiang, G., Cui, Y., Mukherjee, A., Ma, W. L. & Chatterjee, A. K. (1999). $KdgR_{Ecc}$ negatively regulates genes for pectinases, cellulase, protease, $Harpin_{Ecc}$, and a global RNA regulator in *Erwinia carotovora* subsp. *carotovora*. *J Bacteriol* **181**, 2411–2421.

Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mironov, A. A., Koonin, E. V., Roytberg, M. A. & Gelfand, M. S. (1999). Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res* **27**, 2981–2989.

Mironov, A. A., Vinokurova, N. P. & Gel'fand, M. S. (2000). Software for analyzing bacterial genomes. *Mol Biol* **34**, 253–262.

Nasser, W. & Reverchon, S. (2002). H-NS-dependent activation of pectate lyases synthesis in the phytopathogenic bacterium *Erwinia chrysanthemi* is mediated by the PecT repressor. *Mol Microbiol* **43**, 733–748.

Nasser, W., Reverchon, S. & Robert-Baudouy, J. (1992). Purification and functional characterization of the KdgR protein, a major repressor of pectinolysis genes of *Erwinia chrysanthemi*. *Mol Microbiol* **6**, 257–265.

Nasser, W., Awade, A. C., Reverchon, S. & Robert-Baudouy, J. (1993). Pectate lyase from *Bacillus subtilis*: molecular characterization of the gene, and properties of the cloned enzyme. *FEBS Lett* **335**, 319–326.

Nasser, W., Reverchon, S., Condemine, G. & Robert-Baudouy, J. (1994). Specific interactions of *Erwinia chrysanthemi* KdgR repressor with different operators of genes involved in pectinolysis. *J Mol Biol* **236**, 427–440.

Nasser, W., Robert-Baudouy, J. & Reverchon, S. (1997). Antagonistic effect of CRP and KdgR in the transcription control of the *Erwinia chrysanthemi* pectinolysis genes. *Mol Microbiol* **26**, 1071–1082.

Negre, D., Oudot, C., Prost, J. F., Murakami, K., Ishihama, A., Cozzone, A. J. & Cortay, J. C. (1998). FruR-mediated transcriptional activation at the *ppsA* promoter of *Escherichia coli*. *J Mol Biol* **276**, 355–365.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Niersbach, M., Kreuzaler, F., Geerse, R. H., Postma, P. W. & Hirsch, H. J. (1992). Cloning and nucleotide sequence of the *Escherichia coli* K-12 *ppsA* gene, encoding PEP synthase. *Mol Gen Genet* **231**, 332–336.

Nomura, K., Nasser, W. & Tsuyumu, S. (1999). Self-regulation of Pir, a regulatory protein responsible for hyperinduction of pectate lyase in *Erwinia chrysanthemi* EC16. *Mol Plant–Microbe Interact* **5**, 385–390.

Oh, M. K., Rohlin, L., Kao, K. C. & Liao, J. C. (2002). Global expression profiling of acetate-grown *Escherichia coli*. *J Biol Chem* **277**, 13175–13183.

Resibois, A., Colet, M., Faelen, M., Schoonejans, E. & Toussaint, A. (1984). PhiEC2, a new generalised transducing phage of *Erwinia chrysanthemi*. *Virology* **137**, 102–112.

Reverchon, S., Expert, D., Robert-Baudouy, J. & Nasser, W. (1997). The cyclic AMP receptor protein is the main activator of the pectinolysis genes in *Erwinia chrysanthemi*. *J Bacteriol* **179**, 3500–3508.

Robert-Baudouy, J., Nasser, W., Condemine, G., Reverchon, S., Shevchik, V. E. & Hugouvieux-Cotte-Pattat, N. (2000). Regulation of pectinase gene expression in *Erwinia chrysanthemi*. *Plant–Microbe Interact* **5**, 221–268.

Rodionov, D. A., Mironov, A. A., Rakhmaninova, A. B. & Gelfand, M. S. (2000). Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Mol Microbiol* **38**, 673–683.

Roeder, D. L. & Collmer, A. (1985). Marker-exchange mutagenesis of pectate lyase isozyme gene in *Erwinia chrysanthemi*. *J Bacteriol* **164**, 51–56.

Sabnis, N. A., Yang, H. & Romeo, T. (1995). Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. *J Biol Chem* **270**, 29096–29104.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Surgey, N., Robert-Baudouy, J. & Condemine, G. (1996). The *Erwinia chrysanthemi pecT* gene regulates pectinase gene expression. *J Bacteriol* **178**, 1593–1599.

Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**, 33–36.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Thomson, N. R., Nasser, W., McGowan, S., Sebaihia, M. & Salmond, G. P. (1999). *Erwinia carotovora* has two KdgR-like proteins belonging to the IclR family of transcriptional regulators: identification and characterization of the RexZ activator and the KdgR repressor of pathogenesis. *Microbiology* **145**, 1531–1545.

Yew, W. S. & Gerlt, J. A. (2002). Utilization of L-ascorbate by *Escherichia coli* K-12: assignments of functions to products of the *yjf-sga* and *yia-sgb* operons. *J Bacteriol* **184**, 302–306.