Comparative Genomics of Regulation of Fatty Acid and Branched-chain Amino Acid Utilization in Proteobacteria[¶]

Running title: Comparative genomics of FA and ILV degradation

Alexey E. Kazakov[1†], Dmitry A. Rodionov[1,2†*], Eric Alm[3,4], Adam Paul Arkin[4,5,6], Inna Dubchak[4,5], and Mikhail S.Gelfand[1,7]

Institute for Information Transmission Problems (the Kharkevich Institute) RAS, Moscow, Russia,[1] *Burnham Institute for Medical Research, La Jolla, California,*[2] *Departments of Biological Engineering and Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts,*[3] *Virtual Institute of Microbial Stress and Survival, Lawrence Berkeley National Laboratory, Berkeley, California,*[4] *Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California,*[5] *Department of Bioengineering, University of California, Berkeley, California,*[6] *and Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*[7]

[†]A.E.K. and D.A.R. contributed equally to the paper.

*Corresponding author. Mailing address: 10901 North Torrey Pines Road, La Jolla, CA, 92037. Phone: (858) 646-3100. Fax: (858) 795-5249. E-mail: rodionov@burnham.org.

[¶]Supplemental material for this article may be found at http://jb.asm.org/.

Abstract

Bacteria can use branched-chain amino acids (ILV, i.e. isoleucine, leucine, valine) and fatty acids (FA) as sole carbon and energy sources convering ILV into acetyl-CoA, propanoyl-CoA and propionyl-CoA, respectively. In this work, we used the comparative genomic approach to identify candidate transcriptional factors and DNA motifs that control ILV and FA utilization pathways in proteobacteria. The metabolic regulons were characterized based on the identification and comparison of candidate transcription factor binding sites in groups of phylogenetically related genomes. The reconstructed ILV/FA regulatory network demonstrates considerable variability and involves six transcriptional factors from the MerR, TetR and GntR families binding to eleven distinct DNA motifs. The ILV degradation genes in gamma- and beta-proteobacteria are mainly regulated by a novel regulator from the MerR family (e.g., LiuR in *Pseudomonas aeruginosa*) (40 species), in addition, the TetR-type regulator LiuQ was identified in some beta-proteobacteria (8 species). Besides the core set of ILV utilization genes, the LiuR regulon in some lineages is expanded to include genes from other metabolic pathways, such as the glyoxylate shunt and glutamate synthase in the *Shewanella* species. The FA degradation genes are controlled by four regulators including FadR in gamma-proteobacteria (34 species), PsrA in gamma- and beta-proteobacteria (45 species), FadP in beta-proteobacteria (14 species), and LiuR orthologs in alpha-proteobacteria (22 species). The remarkable variability of the regulatory systems associated with the FA degradation pathway is discussed from the functional and evolutionary points of view.

Introduction

Proteobacteria comprise one of the largest divisions within prokaryotes and incorporate species possessing a very complex collection of phenotypic and physiological attributes including many phototrophs, heterotrophs and chemolithotrophs. The proteobacterial group is of great biological significance as it includes a large number of pathogens and symbionts of animals and plants. Thus proteobacteria display an amazing versatility in their ability to use various carbon sources such as carbohydrates, nucleotides, amino acids and lipids. Degradation of branched-chain amino acids valine, leucine, and isoleucine (ILV) and fatty acids (FA) is used for ATP and energy production by many proteobacteria.

The ILV degradation pathways are outlined in Fig. 1A. The first reaction is transamination to the corresponding α-keto acids using either branched-chain amino acid aminotransferase (BCAT) or leucine dehydrogenase (LeuDH). The second step is oxidative decarboxylation to the corresponding acyl-CoA derivative coupled to dehydrogenation, which is carried out by a common branched-chain □-keto acid dehydrogenase (BCDH) complex. Further conversion of branched-chain acyl-CoA derivatives of ILV amino acids, namely isovaleryl-CoA for leucine, 2-methylbutanoyl-CoA for isoleucine, and isobutyryl-CoA for valine, into acetyl-CoA and propionyl-CoA is mediated by individual ILV catabolic pathways (26).

Many enzymes in these downstream ILV degradation pathways belong to large families of paralogs and thus most early annotations of the corresponding genes in bacterial genomes were rather nonspecific “general class” functional assignments. A subsystem-based approach to genome annotations as implemented in the SEED platform (http://theseed.uchicago.edu) was used for reconstruction of the ILV degradation pathways in bacteria (31). A combination of

functional and genome context analysis, as depicted in the SEED Viewer subsystems "Leucine degradation", "Isoleucine degradation" and "Valine degradation" at http://seed-viewer.theseed.org/, provided convincing evidence for the presence of the ILV catabolic pathways in a number of diverse bacteria (31). According to this analysis, the ILV catabolic pathways are present in many lineages of γ-proteobacteria (e.g. in *Pseudomonas aeruginosa*, *Shewanella oneidensis*) with a notable exception of *Escherichia coli* and other enterobacteria.

The *liu* gene cluster involved in leucine and isovalerate utilization was recently identified and characterized in *P. aeruginosa* (1,12,16). Functional roles of the *liuABCDE* genes are shown in Fig. 1A. The first gene encodes a hypothetical transcription factor from the MerR family, called LiuR (12). Although expression analysis of the *liu* genes showed their specific induction by leucine (1), a possible role of LiuR in transcriptional regulation of the *liu* genes has not yet been investigated. The BCDH-encoding operon *bkd* in *Pseudomonas putida* is regulated by an ILV-responsive transcriptional activator BkdR from the AsnC family (25).

The FA degradation pathway is catalyzed by enzymes encoded by the *fad* regulon (see Fig. 1B) (13). Long-chain FAs are transported across the cell membrane using the outer membrane transporter FadL and the inner membrane-associated CoA ligase FadD. After uptake, FAs can be degraded via the β-oxidation pathway and used as an energy/carbon source via the TCA cycle or, alternatively, FAs can be used as precursors for the membrane-phospholipid biosynthesis. The β-oxidative cleavage of acyl-CoAs acts in a cyclic manner and involves their conversion to enoyl-CoAs catalyzed by acyl-CoA dehydrogenase (e.g., FadE), which is followed by hydration, oxidation, and thiolytic cleavage performed by the FA oxidation complex (e.g., FadB-FadA, or FadI-FadJ). In addition, FadH is used for degradation of unsaturated FAs.

Transcriptional control of the FA metabolism in *E. coli* is mediated by the FadR regulatory protein from the GntR family, which recognizes a 17-bp palindromic motif with the consensus sequence AACTGGTCnGACCAGTT (7). FadR senses FA availability in the environment and is released from DNA in the presence of long chain acyl-CoA (17). FadR acts as a repressor of the FA degradation operons (*fadL, fadD, fadE, fadBA, fadH,* and *fadIJ*), and as an activator of the *fabB* and *fabA* genes involved in the unsaturated FA synthesis (5,37). In addition, *E. coli* FadR is involved in the regulatory cascade by activation of the *iclR* gene encoding a repressor of the glyoxylate shunt operon *aceBAK* (15). The comparative genomic analysis of FadR-binding sites demonstrated conservation of the *E. coli* FadR regulon in enterobacteria and its considerable reduction in other γ-proteobacteria, represented by *Haemophilus influenzae* and *Vibrio cholerae* (37). A different transcription factor from the TetR family was recently identified in *Bacillus subtilis* (encoded by the *ysiA* gene) as the master regulator of the *fad* genes. It recognises YsiA-boxes with the consensus TGAATGAnTAnTCATTCA (27). Apart from *E. coli* and *B. subtilis*, the mode of regulation of the FA degradation genes in other bacterial lineages remains unclear.

In this study, we expanded the use of the comparative genomics approach to regulation (for a recent review see (35)) to characterize novel regulons controlling ILV and FA degradation pathways in the α-, β-, and γ-Proteobacteria. The analysis of conserved operons involved in these pathways, initiated in *S. oneidensis* and related bacteria, led to a tentative identification of two novel regulons characterized by unique DNA motifs. These highly conserved motifs are candidate binding sites of two different subfamilies of transcription factors which are widely distributed in proteobacteria. Their representatives in *P. aeruginosa* were previously described as LiuR and PsrA. In particular, the *P. aeruginosa* regulator PsrA from the TetR family was

originally shown to be involved in the stationary phase-induced transcriptional regulation of *rpoS* and some other genes (19,20,21). Here we perform comparative genomic reconstruction of the respective regulons in proteobacteria and report that their major targets are the ILV and FA degradation pathways, respectively. Finally, we identified and characterized non-orthologous regulons for the ILV and FA degradation in the *Burkholderiales* group of β-proteobacteria (named here LiuQ and FadP, respectively). Distribution and partial overlap between three FA degradation regulons (FadR, PsrA and FadP) and two ILV degradation regulons (LiuR and LiuQ) in γ-, and β-proteobacteria and its evolutionary implications are discussed.

Data and Methods

Bacterial genome sequences were downloaded from GenBank (3). The gene identifiers from GenBank are used throughout. Protein similarity search was done using the Smith-Waterman algorithm implemented in the Genome Explorer program (28). Orthologous proteins were initially defined by the best bidirectional hits criterion and if necessary confirmed by the analysis of phylogenetic trees. The phylogenetic trees were constructed by the maximum likelihood method implemented in the PHYLIP package (10) using multiple sequence alignments of protein sequences produced by ClustalX (38).

A simple iterative procedure implemented in the program SignalX (as described previously in (14) and recently reviewed in (35)) was used for construction of transcription factor–binding motifs in sets of upstream fragments of potentially coregulated genes. For the LiuR regulon, the original training set included the ILV degradation operons in *S. oneidensis*. Orthologs of these and other candidate members of the predicted *S. oneidensis* LiuR regulon identified in other γ-proteobacteria (see Table S1 in the supplemental material) were used as a training set for construction of the 'LiuR_gamma' profile. For the PsrA regulon, the training set used for the 'PsrA_gamma' profile construction included the FA degradation operons from *S. oneidensis*, as well as from *Vibrio* and *Pseudomonas* species (see Table S5). The resulting LiuR and PsrA binding site profiles were used for the comparative analysis of the respective regulons in γ-, β-, and α-proteobacteria. For the FadR regulon, we started from the training set of known members of this regulon in *E.coli* and their orthologs in other Enterobacteriales (37). Finally, for each taxonomic group of γ-proteobacteria with the FadR regulon (Enterobacteriales, Vibrionales, Pasteurellales, Altermonadales) we used a separate training set of the upstream regions of

candidate FadR target operons to construct the FadR binding site profile (see Table S4). For the LiuQ and FadP regulons in β-proteobacteria, the training sets included the ILV and FA degradation operons, respectively (see Table S2, S6).

Each genome encoding the studied transcription factor was scanned with the constructed profile using the GenomeExplorer software (28), and genes with candidate regulatory sites in the upstream regions were selected. We analysed only 5'-untranslated gene regions up to 400 nt upstream of the translation start site. z-scores of candidate sites were calculated as the sum of the respective positional nucleotide weights. The threshold for the site search was defined as the lowest score observed in the training set (see Tables S1-S6). The consistency check of the predicted members of regulons was used to eliminate false-positive site predictions. This approach is based on the assumption that regulatory events tend to be conserved in closely related species with orthologous regulators (35). The upstream regions of genes that are orthologous to genes containing conserved regulatory sites were examined for candidate sites even if these were not detected automatically with a given threshold (weak regulatory sites with scores below the threshold are underlined in Tables S1-S6). Among candidate members of PsrA and FadP regulons, only genes having candidate sites conserved in at least two other genomes were retained for further analysis. For the PsrA regulon, we also considered several candidate regulon members that did not satisfy this conservation criterion, but were functionally related to the fatty acid metabolism. Sequence logos for the derived regulatory motifs were drawn using the WebLogo package v.2.6 (8) (http://weblogo.berkeley.edu/).

Results

ILV degradation regulons in proteobacteria

Identification of the LiuR regulon in *S. oneidensis.* Novel regulon for ILV degradation genes was initially identified by analysis of gene expression data in *S. oneidensis*. Using microarray data on salt (23) and alkaline (22) stresses, we selected 15 most up-regulated genes constituting three potential operons, namely SO1898-91, SO1677-83, and SO2339-41. According to the reconstructed metabolic pathways in the SEED database, the above three operons belong to the ILV degradation subsystem (see 'Branched chain amino acid degradation regulons' subsystem available online at http://theseed.uchicago.edu/FIG/subsys.cgi). By applying the motif recognition procedure to a training set of upstream regions of these operons n *S. oneidensis* and orthologous operons in other *Shewanella* species, we found a common 18-bp DNA motif named ILV-box (see Table S1 in the supplemental material). The consensus sequence for this palindromic motif (ILV-box) is sTTTACGTwwACGTAAAs, where 'w' and 's' denote 'A or T' and 'C or G', respectively (see the motif logo in Fig. 2A). We also identified an additional motif gTGTAAAnnnnnntTTACAc of the aromatic amino acid-responsive regulator TyrR that we had studied in detail elsewhere (36). Candidate TyrR-binding sites were observed upstream of three mentioned operons and the SO2638 gene. The TyrR regulon was not further analyzed in this study because it does not include the ILV degradation genes in other bacterial species outside of the *Shewanella* group.

The SO1898-SO1893 genes are orthologs of the *liuRABCDE* genes involved in leucine and isovalerate utilization in *P. aeruginosa* (12) (Fig. 3). The first gene in this cluster encodes a hypothetical transcription factor from the MerR family, named LiuR for a candidate regulator of the *liu* cluster (12). In the genomes of most γ- and β-proteobacteria, the scanning with the ILV-box recognition profile identified candidate regulatory sites upstream of operons containing orthologs of *liuR*. We tentatively attributed the ILV-box motif to the LiuR transcription factor

based on the following comparative genomic evidence: (i) positional clustering on the chromosome of the *liuR* genes and ILV-boxes, and (ii) correlation in the phylogenetic pattern of co-occurrence of *liuR* and ILV-boxes in the genomes of various proteobacteria (see the next section for details).

The second LiuR-regulated operon in *S. oneidensis* (SO1677-1683) encodes all enzymes required for the utilization of 2-methylbutanoyl-CoA and isobutyryl-CoA, the products of the isoleucine and valine degradation, therefore it was named the *ivd* operon (Fig. 1 and Fig. 3B). The third candidate member of the LiuR regulon is the BCDH enzyme complex encoded by the *bkd* operon (SO2339-2341) and involved in the second step of ILV utilization. Scanning the *S. oneidensis* genome with the constructed ILV-box profile identified six more operons that are likely regulated by LiuR (see Table S1 in the supplemental material). These additional candidate members of LiuR regulon include leucine dehydrogenase *ldh*, glyoxylate shunt genes *aceBA*, glutamate synthase *gltBD*, threonine synthesis operon *thrABC*, regulator of aromatic amino acid metabolism *tyrR* and electron transfer flavoprotein (ETF) operon *etfBA*. The latter operon is directly connected to the ILV degradation pathway, as isovaleryl-CoA dehydrogenase (IVD) is known to utilize ETF as an electron acceptor in eukaryotes (11) (see Fig. 1A). Search for similar LiuR binding sites in the genomes of other twelve *Shewanella* species confirmed conservation of the LiuR regulon in the *Shewanella* genus with the only exception being the *tyrR* gene, which has a LiuR binding site only in eight species (See Table S1 in the supplemental material).

LiuR orthologs show mosaic distribution in the genomes of γ-, β-, and α-proteobacteria. There are no orthologs in other taxonomic groups (Table 1). The phylogenetic tree of the LiuR family has three main branches corresponding to the three subdivisions of proteobacteria (Fig.

4A); the respective binding motifs are represented in Fig. 2. Most α-proteobacteria, as well as some β-, and γ-proteobacteria order have two LiuR paralogs. The analysis of the genome context and the reconstruction of the LiuR regulons is outlined below and summarized in Table 1.

Reconstruction of the LiuR regulon in γ-proteobacteria. Orthologs of *liuR* (whose DNA motif is given in Fig. 2A) were found in four lineages of γ-proteobacteria: *Alteromonadales*, *Vibrionales*, *Pseudomonadales*, and *Oceanospirillales*. They are always located in the leucine degradation *liu* gene clusters at the first position (See Table S1 in the supplemental material). Furthermore, in the *Vibrionales* and some *Alteromonadales*, the *liu* operons form a supercluster with the isoleucine/valine degradation *ivd* operons. Genomic identification of LiuR binding sites combined with the comparative regulon consistency check (35) (for details see Materials and Methods) lead to tentative reconstruction of the LiuR regulons in these genomes (See Table S1 in the supplemental material).

As in the *Shewanella* species, the conserved core of the LiuR regulon in four other *Alteromonadales* genomes is represented by the *liu*, *ivd*, *bkd*, and *etf* gene clusters involved in the ILV degradation. However, other predicted members of *Shewanella* LiuR regulon (*aceBA*, *thrABC*, *tyrR* and *gltBD*) are not controlled by LiuR in these species. The *etfBA* genes were found in one LiuR-regulated cluster with the ETF-ubiquinone oxidoreductase *etfD* in the *Pseudoalteromonas* species. Similarly to the *Shewanella* species, candidate LiuR binding sites were found upstream of the *ldh* gene in *Pseudoalteromonas haloplanktis* but this regulatory interaction is not conserved in other *Alteromonadales*. The *liu* operon in *Idiomarina loihiensis* contains an additional gene *IL0879* encoding acetoacetyl-CoA synthase (*aacS*), which is absent from all *Shewanella* species. A strong LiuR site was also detected upstream of the acyl-CoA

dehydrogenase gene PSHAb0374 (called *acdH*) in *P. haloplanktis*, but its orthologs in other *Alteromonadales* are not regulated by LiuR.

The structure of the LiuR regulon in the *Vibrionales* is similar to that of the *Alteromonadales*, although the *ldh* gene is missing in these genomes and the *bkd* operon is not regulated by LiuR. *V. parahaemolyticus* contains two LiuR-regulated copies of the *ivd* and *liu* operons, probably, as a result of a recent duplication; a candidate LiuR binding motif was also found upstream of the seven-gene operon VPA1153-1147 encoding the ABC transporter for branched-chain amino acid LivGHMKF and two hypothetical enzymes.

The LiuR regulon in the *Pseudomonadales* is much smaller and includes the *liu* operon and the acetoacetyl-CoA synthase gene *aacS*. In addition, a candidate LiuR binding site was found upstream of the *etfBA/D* genes only in *Pseudomonas fluorescens*. Although the final reaction in the leucine utilization pathway in *Pseudomonas aeruginosa* and *S. oneidensis* is represented by two different enzymes, AACS and SCOT, respectively (Fig. 1A), both these alternative enzymes belong to the respective LiuR regulons (Fig. 3). The *P. aeruginosa aacS* gene (PA2557) is preceded by a LiuR-binding site, whereas SCOT is encoded by last two genes (*liuFG*) within the LiuR-regulated *liu* operon in *S. oneidensis*.

The LiuR regulon in two *Oceanospirillales* species, *Hahella chejuensis* and *Alcanivorax borkumensis*, includes the *liu-aacS* operon (See Table S1 in the supplemental material). In addition, a candidate LiuR binding site precedes the *bkd* gene cluster in *H. chejuensis.*

Comparative analysis of the LiuR regulons in β-proteobacteria. The DNA motif of the β-proteobacterial LiuR is largely similar to that of γ-proteobacteria (Fig. 2B).

Chromobacterium violaceum has two highly similar *liuR* paralogs likely resulting from a

recent duplication (Table 1). The reconstructed LiuR regulon in this genome is most similar to the γ-proteobacterial ones. It includes various ILV degradation genes organized in the *liu-aacS* operon and *aacS2* paralog that are clustered with *liuR1* regulatory gene, and the *ivd* operon co-localized with *liuR2* (See Table S1 in the supplemental material).

In β-proteobacteria from the *Rhodocyclales* and *Burkholderiales* orders, the composition of the reconstructed LiuR regulons is highly variable and differs significantly from the LiuR regulon in γ-proteobacteria (Table 1). Within these taxonomic groups, the *liuR* gene is always co-localized with operons that include the isovaleryl-CoA dehydrogenase *liuA*, however the size and composition of these candidate LiuR-regulated gene clusters vary from just one gene *liuA* in *Bordetella pertussis* up to 25 genes in *Ralstonia eutropha* (See Table S1 in the supplemental material). Additional genes within the ILV degradation gene clusters include carbonic anhydrase *cah*, isocitrate dehydrogenase phosphatase/kinase *aceK*, biotin biosynthesis genes *bioAFDB*, and many hypothetical genes (e.g. *paaI* and *gloB*). Other candidate members of the LiuR regulons in β-proteobacteria are the ETF operon *etfBA*, methylmalonyl-CoA mutase *mcm*, malate dehydrogenase *mdh*, 3-hydroxyacyl-CoA dehydrogenase *paaH*, and short-chain-specific acyl-CoA dehydrogenase *acdH* (Table 1).

Identification of the LiuQ regulon in *Burkholderiales.* LiuR orthologs were not identified in most *Burkholderia* species and *Methylibium petroleiphilum*. However, the *liu* operons without *liuR* genes are present in these genomes, and another transcriptional regulator from the TetR family, named *liuQ*, was found adjacent to the *liu* operons. By applying the motif recognition procedure to the training set of upstream regions of these *liu* operons, we identified a conserved DNA motif with the palindromic consensus TTGAGynnnrCTCAA, where ‘y’ and ‘r’ denote ‘C

or T ‘and ‘A or G’, respectively (Fig. 2D). Such sites are present in two copies in the common upstream region of the *liuQ* and *liABCD* operons (Table S2). We propose that these palindromes are the binding sites of the LiuQ dimers. The LiuQ and LiuR regulons have overlapping distribution in two *Ralstonia* and one *Burkholderia* species (Table 1). In *R. eutropha*, LiuR and LiuQ regulate different operons that contain paralogous copies of the *liuABDE* genes. In *R. metallidurans*, the *liuAC* genes belong to the LiuR-regulated operon, whereas the *liuBDE* genes are regulated by LiuQ. In *B. xenovorans*, the *liuABCD* operon is under dual regulation of LiuR and LiuQ (Table 1).

Comparative analysis of the LiuR regulons in α-proteobacteria. In *Rhodospirillum rubrum*, the reconstructed LiuR regulon includes the ILV degradation *liu* and *ivd* operons, the *liuR* gene, and an ortholog of the *V. vulnificus* VPA1153-1147 operon encoding the branched-chain amino acid ABC transporter LivGHMKF (Table 1).

Besides *R. rubrum*, the *liuR* orthologs (and additional paralogs in some genomes) are present in twenty α-proteobacteria. The consensus of candidate LiuR-binding sites in these α-proteobacteria is very similar to the LiuR consensus in β-, and γ-proteobacteria (Fig. 2C), however, the composition of the LiuR regulons is completely different and thus they were designated as LiuR$^{\alpha}$ regulons. Tentative metabolic reconstruction suggests that in four groups of α-proteobacteria (*Rhizobiales*, *Caulobacterales*, *Sphingomonadales*, *Rhodobacterales*) the LiuR$^{\alpha}$ regulon controls genes from the fatty acid (FA) degradation pathway, such as the fatty acid oxidation complex *acdAB*, and acyl-CoA dehydrogenases *acdH* and *acdL* (Table 2 and Table S3). In the *Rhizobiales* group, additional candidate LiuR-binding sites were found upstream of genes involved in the FA degradation (*fadD*, *etfAB*, *etfD*, *hbdA*), and the TCA cycle (*mdh*,

sucCDAB, lpdA).

Some α-proteobacteria, such as *Brucella*, *Mesorhizobium* and *Rhodobacter* spp., have two LiuR paralogs (see the phylogenetic tree in Fig. 4A). We were not able to identify systematic differences between the candidate LiuR-binding sites in the genomes harboring these genes. Thus we were not able to assign the regulated operons to either of the two paralogs. These findings are in line with the observed high conservation of the DNA-binding (N-terminal) domains of LiuR from α-proteobacteria and two other groups of proteobacteria and weak similarity of their ligand-binding (C-terminal) domains (data not shown). Probably, LiuR paralogs senses different ligands but can recognize similar if not identical sequence motifs.

Fatty acid (FA) degradation regulons in proteobacteria

Reconstruction of the FadR regulon in γ-proteobacteria. The GntR-like transcriptional factor FadR in *E. coli* is a negative regulator of genes involved in the FA degradation (*fadBA, fadD, fadE, fadIJ*) and transport (*fadL*) and an activator of two genes involved in the unsaturated FA synthesis (*fabA* and *fabB*) and also the *iclR* gene encoding the repressor of the glyoxylate shunt operon *aceBAK* (5, 7, 15, 17, 37). Orthologs of FadR are present in four groups of γ-proteobacteria (*Enterobacteriales*, *Pasteurellales*, *Vibrionales*, and *Alteromonadales*). The recognition profile for FadR-binding sites in enterobacteria was constructed using the set of upstream regions of known FadR-regulated operons in *E. coli* and orthologous operons in other enterobacteria. Genomic searches with the constructed profile demonstrated high conservation of the FadR regulon in Enterobacteriales, whereas the content of the FadR regulon in other groups of γ-proteobacteria differed to some extent (Table S4). A series of taxonomic group-specific

FadR recognition profiles was constructed and used for detailed comparative reconstruction of the FadR regulons in four groups of γ-proteobacteria (Table 3).

A highly conserved core of the FadR regulon in the *Enterobacteriales* is formed by the FA degradation genes (Table 3). In contrast, FadR binding sites upstream of the FA synthesis, transport and the *iclR* genes are not strictly conserved in the enterobacteria. The FadR-IclR regulatory cascade, where FadR negatively regulates the *aceBAK* operon by activation of the IclR repressor, is conserved only in genomes closely related to *E. coli* such as the *Shigella* and *Salmonella* species. Interestingly, the *aceBAK* operon in the *Yersinia* species is preceded by a candidate FadR-binding site suggesting a rewiring of the regulatory cascade (Table S4).

In the *Vibrionales*, the conserved core of the FadR regulon is formed by the FA degradation operons *fadBA*, *fadE*, and *fadIJ*, as well as the phospholipid synthesis gene *plsB*, whereas the FadR-dependent regulation of *fadH* and *fadL* is not conserved in two or more species (Table 3). In the *Pasteurellales*, most FA degradation genes are absent, and the FadR regulon contains only genes involved in the FA transport (*fadL*) and synthesis (*fabA*, *fabB*, *fabDG*, *fabI*, and *accA*). In the *Alteromonadales*, the FA degradation operon *fadIJ* is the only conserved member of the FadR regulon. In the *Pseudoalteromonas* species, there are two additional FadR targets, the *fadBA* and *fadH* operons. In most *Shewanella* species, the FadR regulon also includes the FA transporter *fadL*, and two hypothetical genes encoding probable enoyl-CoA hydratase (SO0572) and Acyl-CoA N-acyltransferase (SO4716).

Identification of the PsrA regulon in proteobacteria. The FadR regulon was not found in several groups of γ-proteobacteria (e.g., *Pseudomonadales*, *Xanthomonadales*, *Oceanospirillales*), whereas in the *Alteromonadales* group (e.g., in the *Shewanella* species) it

includes only a small subset of FA degradation (FAD) genes. In an attempt to identify a novel FAD regulatory system, we applied the motif detection procedure to the upstream regions of the *fad* genes from thirteen *Shewanella* species and identified a 20-bp palindromic motif (named FAD-box, Fig. 2E). Genomic searches with the FAD-box profile followed by the inter-genomic consistency check allowed us to tentatively reconstruct the novel FAD regulon in the *Shewanella* species. This regulon includes most FA degradation genes, as well as the glyoxylate shunt *aceBA* operon, the TCA cycle *sdh* operon and several hypothetical genes (Table 3). A candidate regulatory gene that encodes a TetR-type transcriptional factor (named PsrA, by the name of its ortholog previously characterized in *Pseudomonas aeruginosa*) is preceded by a FAD-box and forms a putative operon with the *fadE* gene in *Alteromonadales* and *Vibrionales* (Table S5). Orthologs of the *psrA* gene preceded by candidate FAD-boxes were identified in other groups of the γ-proteobacteria (e.g. *Pseudomonadales*), as well as in various groups of the β- and α-proteobacteria (Fig. 2F), and in many genomes they are co-localized with FA degradation genes (Fig. 4B). The phyletic distribution and genomic co-localization of FAD-boxes and *psrA* genes strongly suggest that PsrA is a regulator that recognizes a FAD-box and regulates FA degradation genes in proteobacteria. An overview of the PsrA regulon reconstructed by the comparative genomics approach in proteobacteria is given below.

The FA degradation genes *fadBA*, *fadD*, *fadE*, *fadIJ*, *fadH* and *acdH* are the most conserved members of the PsrA regulon in γ-proteobacteria (Table 3). The ETF operon *etfBA* and the ETF-ubiquinone oxidoreductase gene *etfD* were found within the PsrA regulon in the *Alteromonadales*, *Pseudomonadales*, and *Oceanospirillales*. The *scp* gene encoding a putative sterol carrier protein (COG3255) is an additional member of the PsrA regulon in the

Pseudomonadales and *Oceanospirillales*. A different composition of the PsrA regulon was found in the *Xanthomonadales*, where *psrA* forms a regulated operon with the FA oxidation genes *acdBA*. The additional regulon members are the FA synthesis genes *fabBA* and *accBC* (Table S5). The reconstructed PsrA regulons in several lineages are extended by different sets of genes that are either hypothetical genes, or genes not directly involved in the FA metabolism, e.g., the transcriptional regulator *algQ* in the *Pseudomonas* species, the TCA cycle genes *aceBA* and *sdhCBA* in the *Shewanella* species, and *mdh* in *Vibrio* species.

Interestingly, the PsrA regulator has been previously described in *P. aeruginosa* as the regulator that controls expression of the alternative sigma factor gene *rpoS* (19). Experimentally determined PsrA binding sites in the promoter regions of the PsrA-repressed genes *psrA*, PA0506 (*acdH*), and PA2952 (*etfB*) coincide with the predicted PsrA binding sites (21). For the *rpoS* gene, which is positively regulated by PsrA at the stationary phase, the experimentally determined PsrA binding site is located 411 bp upstream of its translational start point (20) and thus was missed by our procedure.

In the *Pseudomonas* genomes, we performed an additional search with relaxed threshold for the FAD-box profile and identified candidate PsrA binding sites with scores between 4.44 and 4.84 upstream of the *rpoS* genes (Table S5). At that, candidate PsrA site in *P. aeruginosa* coincides with the experimentally identified PsrA site (20). Recently, PsrA was shown to bind to the *fadBA* (PA3014-PA3013) operon promoter region (18). The transcriptome analysis also revealed PsrA-dependent repression of *acdH, etfBA, etf, psrA,* PA1830 and other genes (18), for which PsrA binding sites were identified here (Table S5).

Among β-proteobacteria, the *psrA* gene was found in three lineages, *Rhodocyclales*,

Burkholderiales, and the *Chromobacterium* group (Table 3). *Chromobacterium violaceum* has the largest PsrA regulon, which includes seven operons involved in the FA utilization (*psrA-fadE*, *fadD*, *fadL*, *acdBA*, *acdH*, *etfBA-acdH2*, and *etfD*), three operons involved in the FA biosynthesis (*fabK*, *fabF*, and *aroQ-accBC*), and the *mdh* gene from the TCA cycle. In the *Rhodocyclales* and the *Bordetella* species, the reconstructed PsrA regulon consists of the FA degradation gene cluster containing the *psrA*, *fadE*, *fadAB*, *fadL*, *fadD*, and *acdBA* genes. The PsrA regulon in the *Burkholderia* species includes the *psrA-fadD* gene cluster (in *B. xenovorans*, this gene cluster includes two additional FA degradation genes *fadE* and *fadA*) and the fatty acid biosynthesis gene cluster *fabH-fabD-fabG-acpP-fabF*. In contrast to these β-proteobacteria, the *psrA* gene and *fabH-fabD-fabG-acpP-fabF* gene clusters in the *Ralstonia* species are preceded by putative PsrA binding sites with score just below the threshold.

The only three α-proteobacterial genomes that encode a PsrA ortholog are *Caulobacter crescentus*, *Bradyrhizobium japonicum*, and *Rhodopseudomonas palustris*. In all three genomes, candidate PsrA binding sites were identified upstream of the acyl-CoA dehydrogenase *acdH* gene (Table S5). In *B. japonicum* and *R. palustris*, PsrA binding sites were also found upstream of *psrA* gene itself.

The FadP regulon of the Burkholderiales. The above analysis left a gap in the regulation of the FA utilization genes in most species from the *Burkholderiales* lineage. To fill this gap, we performed additional search for potential regulatory motifs in upstream regions of the FA degradation genes that are not regulated by PsrA. A conserved 16 bp palindromic motif (Fig. 2G) was identified upstream of the gene cluster containing the *acdH*, *acdBA*, and *echH* genes in three *Ralstonia* species, five *Burkholderia* species, and in *Polaromonas* sp. JS666, *Methylibium*

petroleiphilum, and *Rhodoferax ferrireducens* . The first gene in this gene cluster (RSc0472 in *R. solanacearum*) encodes a TetR-like transcriptional regulator (named FadP), which was tentatively proposed to recognize the newly identified FAD regulatory motif in the *Burkholderiales*. Genomic search for similar candidate FadP binding sites in these species identified 4-15 sites per genome (Table S6). Most candidate FadP-regulated genes encode enzymes involved in the FA degradation or metabolism (see Table 3 for details).

In addition to the above described eleven species, FadP orthologs were found in three *Bordetella* genomes. However, only two candidate FadP binding sites were found in each of these genomes: a site upstream of the *fadP* gene implicating its autoregulation, and a site upstream of the BP3678 gene encoding an uncharacterized exported protein.

Discussion

We performed the comparative genomic reconstruction of the transcriptional regulatory network for genes involved in the ILV and FA degradation in the γ- and β-proteobacteria (Fig. 5) and FA degradation in the α-proteobacteria. For ILV utilization genes we report identification of a novel regulator from the MerR family (LiuR) and its DNA recognition motif (ILV-box). The FA degradation genes in *Enterobacteriales* are regulated by the FadR repressor from the GntR family. Here we report identification of a novel transcriptional factor from the TetR family (PsrA) and its conserved motif (FAD-box) that control FA degradation genes in other lineages of γ-proteobacteria and in β-proteobacteria. In addition to these major transcription factors, two novel TetR-like regulators were predicted to control the ILV and FA degradation regulons in the

Burkholderiales group (FadP and LiuQ, respectively) using other DNA motifs. Finally, we report that the LiuR orthologs in α-proteobacteria regulate the FA degradation and TCA cycle genes.

Gene content of the predicted LiuR regulons in γ- and β-proteobacteria is considerably variable (Table 1). The identified core of the LiuR regulon includes the *liu* and *ivd* genes required for conversion of CoA ethers of branched-chain carboxylic acids into CoA ethers of linear chain carboxylic acids for their subsequent utilization through the TCA cycle (Fig. 1A). Though physiological effector molecule for the LiuR regulator is unknown, we propose that one or several intermediates of the ILV degradation pathway, e.g. the CoA ethers of branched-chain carboxylic acids, might be involved in the modulation of LiuR activity.

Some of the ILV degradation genes, such as *aacS, bkd, ldh, etfBA* and *etfD,* are candidate targets of LiuR regulation only in a fraction of the considered genomes. Thus, the complete leucine degradation pathway is regulated by LiuR only in the *Shewanella* species. Interestingly, the *bkd* operon in *P. aeruginosa* is regulated by the ILV-responsive transcriptional activator BkdR (25). The LiuR regulon in the *Shewanella* lineage includes genes that are involved in glutamate synthesis (*gltBA*), glyoxylate shunt (*aceBA*), and threonine biosynthesis (*thrABC*). This apparent extension of the LiuR regulon could be explained by a metabolic connection between the LiuR-controlled metabolic pathways via acetyl-CoA, a final product of the ILV degradation, which is could be utilized for the amino acid biosynthesis pathways via the TCA cycle. In β-proteobacteria, the LiuR regulons are extended by a number of genes with unclear functional roles in the ILV degradation (e.g. *cah, aceK, paaH, gloB, paaI*), as well as the *mdh* and *sdh* genes encoding TCA cycle enzymes, and the *bio* genes involved in the biotin biosynthesis. The latter observation is in line with the role of biotin as a cofactor of

methylcrotonyl-CoA carboxylase (LiuBD).

Finally, among α-proteobacteria, LiuR was found to control the ILV degradation genes only in *R. rubrum*. In contrast, orthologs of LiuR in other α-proteobacteria were predicted to control genes involved in the FA degradation and other pathways (Table 2). The changed content of the $LiuR^{\alpha}$ regulon suggests that its physiological effector might be an acyl-CoA intermediate of the FA degradation pathway.

As seen from transcriptome analyses (22, 23), most of operons regulated by LiuR in *S. oneidensis* were significantly induced by salt and/or alkaline stresses. Salt stress response and branched-chain amino acids metabolism seem to be linked in *Shewanella* species. Firstly, leucine was shown to be an important source of branched-chain FA in cell membrane as growth of *S. gelidimarina* on leucine as a sole carbon source resulted in two-fold increase of the branched-chain FA fraction in te membrane compared with growth on serine or alanine (29). Secondly, concentration of branched-chain FA in *S. gelidimarina* was highly regulated by salt stress conditions, resulted in decrease of branched-chain FA content at high salinity (30). So, LiuR-dependent derepression of ILV degradation genes in *Shewanella* species reduce the pool of branched-chain acyl-CoA thioesthers (starter units for the biosynthesis of branched-chain FA) and resulted in decrease of branched-chain FA proportion in membrane thus regulating the membrane fluidity in salt stress conditions.

Unlike most members of the MerR family, LiuR seems to act solely as a repressor. Indeed, promoters activated by MerR-type transcription factors have an extended (19-20 bp) spacer between the –35 and –10 promoter boxes which contains the transcription factor binding site partially overlapping the –35 element (reviewed in (4)). On the other hand, MerR represses (but

not activates) promoters having standard 17 bp spacers (32, 33)). We did not observe candidate promoters with extended spacers for LiuR-regulated operons, whereas in most cases canonical candidate promoters overlapping the LiuR-binding site could be identified (data not shown).

The fatty acid degradation genes in the *Enterobacteriales* are regulated by the transcriptional regulatory protein FadR (9). We performed a comparative genomic reconstruction of the FadR regulon in other taxonomic groups of γ-proteobacteria and found that despite overall conservation of the FadR-binding motif (Fig. 2H-2J), the regulon composition demonstrated substantial differences (Table 3). We noted that several *fad* genes in the *Vibrionales* and many *fad* genes in the *Alteromonadales* species lack candidate FadR operator sites, suggesting that other factors might be involved in the regulation of these genes.

Using the comparative genomics procedure, we identified the transcriptional factor PsrA as the master regulator of the FA degradation genes in five taxonomic groups of γ-proteobacteria (*Alteromonadales*, *Vibrionales*, *Xanthomonadales*, *Pseudomonadales*, *Oceanospirillales*) and several species of β-proteobacteria (Table 3). The PsrA regulons in different taxonomic groups also demonstrated their significant variability. For example, the FA biosynthesis genes are regulated by PsrA in the *Xanthomonadales* and several β-proteobacteria, whereas the PsrA regulon in the *Pseudomonadales* includes a variety of cellular processes. Since FadR and PsrA co-occur in the *Alteromonadales* and *Vibrionales* species, in some genomes we observed an overlap between two regulons. For instance, the *fadH*, *fadBA* and *fadIJ* operons in the *Vibrionales*; the *fadIJ* operon in *Shewanella* are co-regulated by both FadR and PsrA. In the β-proteobacteria (where only two or three PsrA sites per genome were found), a novel transcriptional factor (FadP) was found to substitute the PsrA function in most of the

Burkholderiales species. Finally, the FA degradation genes in α-proteobacteria were found to be under candidate regulation of LiuR$^{\alpha}$. In summary, this work revealed a large diversity in the transcriptional factors controlling FA degradation pathways in proteobacteria.

An interesting overlap between the LiuR and PsrA regulons was observed in the *Shewanella* species, where candidate binding motifs of both transcription factors were identified in the *aceBA* and *etfBA* regulatory regions. The former encodes for the acetyl-CoA utilization genes, and thus this observation could be explained by the fact that acetyl-CoA is a common product of both FA and ILV degradation pathways. The latter operon (*etfBA*) encodes electron-transfer flavoprotein, which is used as an electron acceptor by LiuR-regulated dehydrogenases of ILV degradation pathway, as well as PsrA-regulated acyl-CoA dehydrogenase. Similarly, the *etfAB/D* genes were found under overlapping regulation of LiuR and FadP in *Polaromonas* and *Rhodoferax* species.

We also observed rewiring of regulatory cascades. Indeed, the cascade FadR→*iclR* + IclR→*aceBAK* of *E. coli* and *Salmonella* spp. corresponds to a streamlined interaction FadR→*aceBAK* in the Yersinia spp., whereas the *aceBA* genes in the *Shewanella* spp are controlled by the novel regulator PsrA. Three feed-forward loops LiuR→*tyrR* + TyrR→(*liu*, *ivd*, *bkd*) + LiuR→(*liu*, *ivd*, *bkd*) are present in eight of twelve *Shewanella* spp., whereas in the remaining *Shewanella* spp. the *tyrR* gene is not regulated by LiuR.

The reconstructed regulatory network suggests that the LiuR and PsrA regulons are the most widespread regulators of the ILV and FA degradation genes in γ- and β-proteobacteria (Fig. 5), respectively. In contrast, the LiuQ and FadP regulons have the narrowest phylogenetic distribution, being identified only in the *Burkholderiales* order. The FadR regulon was identified

in four taxonomic orders of γ-proteibacteria, namely *Enterobacteriales*, *Pasteurellales*, *Vibrionales* and *Alteromonadales*, where it acts either with PsrA (e.g., in *Vibrionales*) or alone (e.g., in *Enterobacteriales*) to control the FA degradation genes. Based on these observations we suggest the most parsimonious evolutionary scenario for the ILV and FA regulons is as follows. LiuQ and PsrA were likely present in the common ancestor of γ-, and β-proteobacteria, they have been partially or fully substituted by LiuQ and FadP in *Burkholdreiales*, and by FadR in some groups of γ-proteobacteria.

The results of this comparative genomics study demonstrate significant variability in the design and composition of the regulatory networks for control of genes from central metabolic pathways. Similar extreme flexibility of transcriptional regulatory networks across various taxonomic groups of bacteria was reported in previous studies (2, 24, 34). The well-characterized FadR regulon that served as a prototype regulon for FA degradation not only underwent many changes by itself, but even may be not the ancestral one. Although the overall picture for the core of the FA and ILV degradation regulons seems to be rather consistent (Fig. 5), many additional members of these regulons were identified (Tables 1-3), whose current functional annotations do not allow to attribute them to the ILV/FA catabolic pathways.

The reconstructed regulatory network needs to be integrated with functionally related networks, and few remaining gaps, such as the regulation of ILV degradation in the *Vibrionales* and α-proteobacteria need to be filled. And of course, although we are convinced that most computationally identified regulatory interactions reported here are real, each particular prediction requires experimental validation.

Acknowledgements

This study was supported by grants from the Howard Hughes Medical Institute (55005610), the Russian Academy of Science (Program "Molecular and Cellular Biology"), and the Russian Foundation for Basic Research (08-04-01000-a).

This work was part of the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The work was initiated during A.K.'s visit to the Lawrence Berkeley National Laboratory.

References

1. **Aguilar, J.A., A. N. Zavala, C. Diaz-Perez, C. Cervantes, A. L. Diaz-Perez, and J. Campos-Garcia.** 2006. The *atu* and *liu* clusters are involved in the catabolic pathways for acyclic monoterpenes and leucine in *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. **72:**2070-2079.

2. **Babu, M. M., N. M. Luscombe, L. Aravind, M. Gerstein and S. A. Teichmann.** 2004. Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. **14:**283-291.

3. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler.** 2007. GenBank. Nucleic Acids Res. **36:**D25-30.

4. **Brown, N. L., J. V. Stoyanov, S. P. Kidd, and J. L. Hobman.** 2003. The MerR family of transcriptional regulators. FEMS Microbiol. Rev. **27:**145-163.

5. **Campbell, J. W., and J. E. Cronan Jr.** 2002. The enigmatic *Escherichia coli fadE* gene is *yafH*. J. Bacteriol. **184:**3759-3764.

6. **Chin-A-Woeng, T. F., D. van den Broek, B. J. Lugtenberg, and G. V. Bloemberg.** 2005. The *Pseudomonas chlororaphis* PCL1391 sigma regulator *psrA* represses the production of the antifungal metabolite phenazine-1-carboxamide. Mol. Plant Microbe Interact. **18:**244-253.

7. **Cronan, J. E. Jr., and S. Subrahmanyam.** 1998. FadR, transcriptional co-ordination of metabolic expediency. Mol. Microbiol. **29:**937-943.

8. **Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner.** 2004. WebLogo: A sequence logo generator. Genome Research. **14:**1188-1190.

9. **DiRusso, C. C., and W. D. Nunn.** 1985. Cloning and characterization of a gene (*fadR*) involved in regulation of fatty acid metabolism in *Escherichia coli*. J. Bacteriol. **161:** 583-588.

10. **Felsenstein, J.** 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics. **5:**164-166.

11. **Finocchiaro, G., M. Ito and K. Tanaka.** 1987. Purification and properties of short chain acyl-CoA, medium chain acyl-CoA, and isovaleryl-CoA dehydrogenases from human liver. J. Biol. Chem. **262:**7982-7989.

12. **Förster-Fromme, K., B. Höschle, C. Mack, M. Bott, W. Armbruster, and D. Jendrossek.** 2006. Identification of genes and proteins necessary for catabolism of acyclic terpenes and leucine/isovalerate in *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. **72:**4819-4828.

13. **Fujita, Y., H. Matsuoka, and K. Hirooka.** 2007. Regulation of fatty acid metabolism in bacteria. Mol. Microbiol. **66:**829-39.

14. **Gelfand, M. S., E. V. Koonin, and A. A. Mironov.** 2000. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. Nucleic Acids Res. **28:**695-705.

15. **Gui, L., A. Sunnarborg, and D. C. LaPorte.** 1996 Regulated expression of a repressor protein: FadR activates *iclR*. J. Bacteriol. **178:**4704-9.

16. **Höschle, B, V. Gnau, and D. Jendrossek.** 2005. Methylcrotonyl-CoA and geranyl-CoA carboxylases are involved in leucine/isovalerate utilization (Liu) and acyclic terpene utilization (Atu), and are encoded by *liuB/liuD* and *atuC/atuF*, in *Pseudomonas aeruginosa*. Microbiology. **151:**3649-3656.

17. **Iram, S. H., and J. E. Cronan.** 2005. Unexpected functional diversity among FadR fatty acid transcriptional regulatory proteins. J. Biol. Chem. **280:**32148-32156.

18. **Kang, Y., D. T. Nguyen, M. S. Son, and T. T. Hoang.** 2008. The *Pseudomonas aeruginosa* PsrA responds to long-chain fatty acid signals to regulate the *fadBA5* beta-oxidation operon. Microbiology. **154:**1584-1598

19. **Kojic, M., and V. Venturi.** 2001. Regulation of *rpoS* gene expression in *Pseudomonas*: involvement of a TetR family regulator. J. Bacteriol. **183:**3712-3720.

20. **Kojic, M., C. Aguilar, and V. Venturi.** 2002. TetR family member PsrA directly binds the *Pseudomonas rpoS* and *psrA* promoters. J. Bacteriol. **184:**2324-2330.

21. **Kojic, M., B. Jovcic, A. Vindigni, F. Odreman, and V. Venturi.** 2005. Novel target genes of PsrA transcriptional regulator of *Pseudomonas aeruginosa*. FEMS Microbiol. Lett. **246:**175-181.

22. **Leaphart, A.B., D. K. Thompson, K. Huang, E. Alm, X. F. Wan, A. Arkin, S. D. Brown, L. Wu, T. Yan, X. Liu, G. S. Wickham, and J. Zhou.** 2006. Transcriptome profiling of *Shewanella oneidensis* gene expression following exposure to acidic and alkaline pH. J. Bacteriol. **188:**1633-1642.

23. **Liu, Y., W. Gao, Y. Wang, L. Wu, X. Liu, T. Yan, E. Alm, A. Arkin, D. K. Thompson, M. W. Fields, and J. Zhou.** 2005. Transcriptome analysis of *Shewanella oneidensis* MR-1 in response to elevated salt conditions. J. Bacteriol. **187:**2501-2507.

24. **Lozada-Chávez, I., S. C. Janga and J. Collado-Vides.** 2006. Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res. **34:**3434-3445.

25. **Madhusudhan, K.T., J. Luo, and J. R. Sokatch.** 1999. In vitro transcriptional studies of the *bkd* operon of *Pseudomonas putida*: L-branched-chain amino acids and D-leucine are the inducers. J. Bacteriol. **181:**2889-94.

26. **Martin, R. R., V. D. Marshall, J. R. Sokatch, and L. Unger.** 1973. Common enzymes of branched-chain amino acid catabolism in *Pseudomonas putida*. J. Bacteriol. **115:**198-204.

27. **Matsuoka, H., K. Hirooka, and Y. Fujita.** 2007. Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. J. Biol. Chem. **282:**5180-5194.

28. **Mironov, A. A., N. P. Vinokurova, and M. S. Gelfand.** 2000. Software for analysis of complete bacterial genomes. Mol. Biol. **34:**222-231.

29. **Nichols, D. S., P. D. Nichols, N. J. Russell, N. W. Davies and T. A. McMeekin.** 1997. Polyunsaturated fatty acids in the psychrophilic bacterium *Shewanella gelidimarina* ACAM 456T: molecular species analysis of major phospholipids and biosynthesis of eicosapentaenoic acid. Biochim. Biophys. Acta. **1347:**164-176.

30. **Nichols, D. S., J. Olley, H. Garda, R. R. Brenner, and T. A. McMeekin.** 2000. Effect of temperature and salinity stress on growth and lipid composition of Shewanella gelidimarina. Appl. Environ. Microbiol. **66:**2422-2429.

31. **Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon , V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Rückert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein.** 2005. The subsystems approach

to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. **33:**5691-5702.

32. **Park, S. J., J. Wireman, and A. O. Summers.** 1992. Genetic analysis of the Tn21 *mer* operator-promoter. J. Bacteriol. **174:**2160-2171.

33. **Parkhill, J., and N. L. Brown.** 1990. Site-specific insertion and deletion mutants in the *mer* promoter-operator region of Tn501; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR. Nucleic Acids Res. **18:**5157-5162.

34. **Price, M. N., P. S. Dehal and A. P. Arkin.** 2007. Orthologous transcription factors in bacteria have different functions and regulate different genes. PLoS Comput. Biol. **3:**1739-1750.

35. **Rodionov, D. A.** 2007 Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem. Rev. **107:**3467-3497.

36. **Rodionov, D. A., M. S. Gelfand, M. Romine, and A. O. Osterman.** 2007. Comparative genomics of transcriptional regulation of metabolic pathways in *Shewanella* species. Abstr. Genomics:GTL Awardee Workshop V, abst. 100.

37. **Sadovskaya, N. S., O. N. Laikova, A. A. Mironov, and M. S. Gelfand.** 2001. Study of regulation of long-chain fatty acid metabolism using computer analysis of complete bacterial genomes. Mol. Biol. **35:**862-866.

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

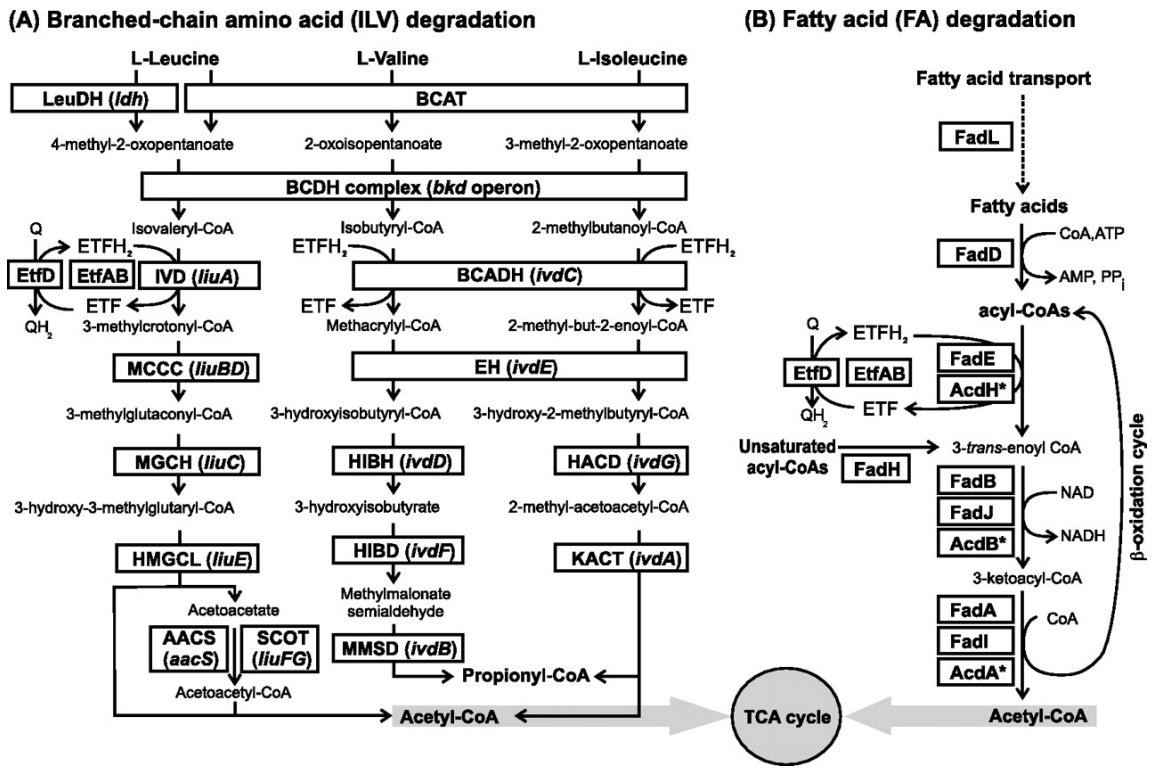

FIG. 1. Branched-chain amino acid (A) and fatty acid (B) catabolic pathways.

(A) Abbreviations of functional roles of ILV catabolic enzymes were adopted from respective SEED subsystems (http://seed-viewer.theseed.org/). LeuDH, leucine dehydrogenase (EC 1.4.1.9); BCAT, branched-chain amino acid aminotransferase (EC 2.6.1.42); BCDH, branched-chain alpha-keto acid dehydrogenase (EC 1.2.4.4); IVD, isovaleryl-CoA dehydrogenase (EC 1.3.99.10); BCADH, branched-chain acyl-CoA dehydrogenase (EC 1.3.99.12); MCCC, methylcrotonyl-CoA carboxylase (EC 6.4.1.4); EH, enoyl-CoA hydratase (EC 4.2.1.17); MGCH, methylglutaconyl-CoA hydratase (EC 4.2.1.18); HIBH, 3-hydroxyisobutyryl-CoA hydrolase (EC 3.1.2.4) ; HACD, 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35); HMGCL, hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4); HIBD, 3-hydroxyisobutyrate dehydrogenase (EC 1.1.1.31); KACT, 3-ketoacyl-CoA thiolase (EC 2.3.1.16); MMSD, methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27). The conversion of acetoacetate to acetoacetyl-CoA is mediated by either of two alternative routes: SCOT, succinyl-CoA:3-ketoacid-CoA transferase (EC 2.8.3.5); AACS, acetoacetyl-CoA synthetase (EC 6.2.1.16). Assimilation of acetoacetyl-CoA and propionyl-CoA products occurs using central metabolic enzymes that are shared with other pathways (not shown).

(B) Fatty acid degradation proteins are abbreviated according to the corresponding gene names in *E. coli*

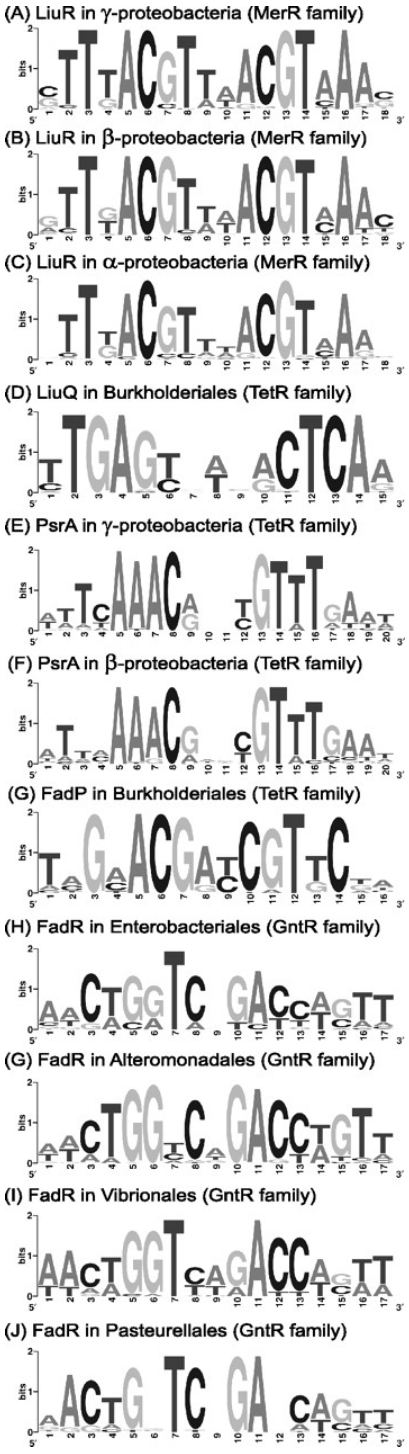

FIG. 2. Sequence logos for the identified binding sites of LiuR, LiuQ, PsrA, FadR, and FadP regulators in various taxonomic groups of Proteobacteria.

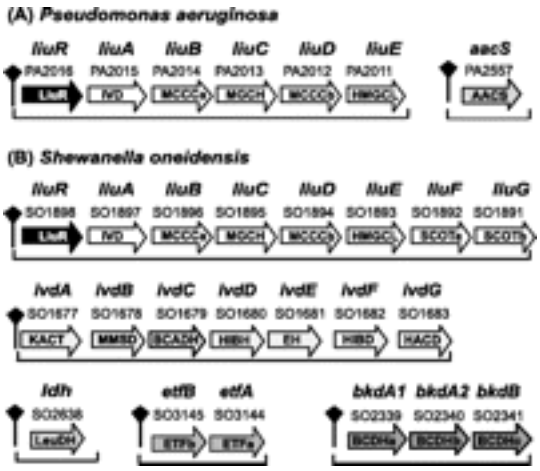

FIG. 3. Branched-chain amino acid degradation operons under control of LiuR regulon in *Pseudomonas aeruginosa* (A), and *Shewanella oneidensis* (B).

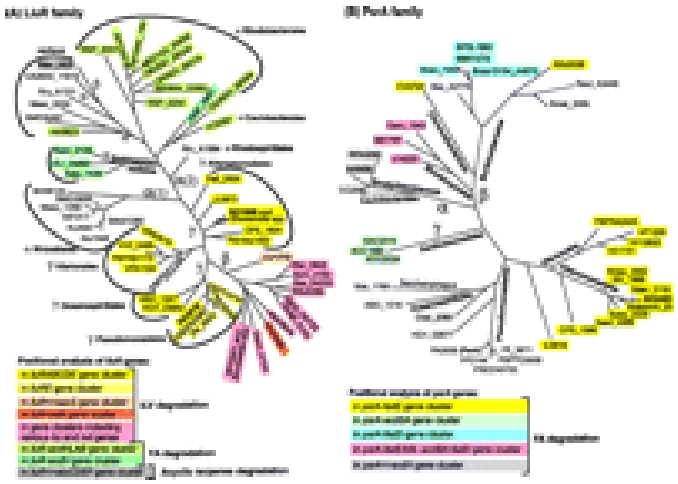

FIG. 4. Maximum likelihood phylogenetic tree of LiuR (A) and PsrA (B) regulators in various groups of proteobacteria. Background colors denote positional linkages of the regulatory genes.

FIG. 5. Candidate regulatory elements and genes involved in the ILV and FA degradation in γ-, and β-proteobacteria. Only conserved members of the predicted ILV/FA regulons are shown. Genes are arranged by the metabolic pathway. Genomes are arranged by taxonomic lineages. When the gene is present in the genome, the background colors denote the presence of the specific recognition motif in its upstream region. If the gene is preceded by two different regulatory motifs, it is shown bi a diagonally separated bicolor square. Genes without any candidate regulatory motifs described in this study are indicated by they grey background color. Empty crossings denote the absence of an orthologous gene in the genome.

TABLE 1. Reconstructed ILV degradation regulons in proteobacteria.

Taxonomic group[1] / organism	**LiuR / LiuQ[2]**	**ILV degradation genes[3]**	**Other co-regulated genes[3]**
γ: Alteromonadales			
Shewanella spp. (13 genomes)	R	*liuABCDEFG; ivdABCDEFG; bkd; ldh; etfBA*	*aceBA; gltBA; thrABC; tyrR (8)*
Idiomarina loihiensis	R	*liuABCDE; ivdA; bkd; aacS*	
Colwellia psychrerythraea	R	*liuABCDEFG; ivdAEG; bkd; etfBA/D*	
Pseudoalteromonas haloplanktis	R	*liuABCDEFG; ivdABCDEFG; bkd; ldh; etfBA/D*	*acdH*
Pseudoalteromonas atlantica	R	*liuABCDEFG; ivdABCDEFG; etfBA/D*	
γ: Vibrionales (3 genomes)	R,R	*liuABCDE; ivdABCDEFG; etfBA/D*	*livGHMKF (1)*
γ: Pseudomonadales (5 genomes)	R	*liuABCDE; aacS (4); etfBA/D (1)*	
γ: Oceanospirillales			
Hahella chejuensis	R	*liuABCDE; bkd; aacS*	
Alcanivorax borkumensis	R	*liuABCD; aacS*	
β: Chromobacterium (1 genome)	R,R	*liuABCDE; ivdABCDEFG; aacS*	
β: Rhodocyclales			
Dechloromonas aromatica	R	*liuABCDE; ivdAC*	*cah; paaI, gloB*
Azoarcus sp.	R	*liuAC; ivdA*	*cah; paaH, gloB*
β: Burkholderiales			
Bordetella spp. (3 genomes)	R	*liuABD; ivdACDF*	*paaI; aceK (2); sdh/glt; mdh*
Ralstonia solanacearum	R	*liuABCDE; ivdAC*	*cah; paaH; aceK; mcm, gloB*
Ralstonia eutropha	R,Q	*liuABDE-1; ivdAC; liuABCDE-2; aacS*	*paaI, cah; paaH; aceK; bio, gloB*
Ralstonia metallidurans	R,Q	*liuAC; ivdAC; liuBDE; aacS*	*cah; paaH; aceK; mcm*
Burkholderia xenovorans	R,Q	*liuABCD; ivdAC; aacS*	*cah; aceK*
Burkholderia spp. (4 genomes)	Q	*liuABCD*	
Methylibium petroleiphilum	Q	*liuABCD*	
Polaromonas sp.	R	*liuA; ivdAC; etfBA*	*paaI, cah; paaH; aceK; mcm; acdH*
Rhodoferax ferrireducens	R,R	*liuA; ivdAC; etfBA*	*cah; paaH; aceK; mcm; acdH*
α: Rhodospirillales (1 genome)	R	*liuRABCDE; ivdACEF*	*livGHMKF*

TABLE 2. Reconstructed FA degradation LiuR$^{\alpha}$ regulon in α-proteobacteria.

Taxonomic group[1] / organism	**FA degradation genes[2]**	**Other co-regulated genes[2]**
α: Rhizobiales		
Rhizobiaceae (4 genomes)	*acdABH; fadD; etfBA; etfD; hbdA; echH*	*mdh-suc-lpdA;* RL0537*;* RL1312*;* mll7718 *(3)*
Brucella melitensis	*acdABH; fadD; etfBA; etfD; hbdA; echH*	*mdh-suc-lpdA;* RL0537*;* RL1312*;* mll7718*;* BMEI0304
Mesorhizobium loti	*acdABH; acdL; echH*	*mdh-suc-lpdA*
Mesorhizobium sp.	*acdABH; fadD; etfBA; hbdA*	*mdh-suc-lpdA*
Bartonella spp. (2 genomes)	-	*mdh-suc-lpdA*
α: Caulobacterales (1 genome)	*acdABH; acdL*	
α: Sphingomonadales		
Novosphingobium aromaticivorans	*acdH; etfBA; acdQP*	Saro_0871-67
Erythrobacter litoralis	*acdA; acdH; etfBA; acdQP*	*maoC*
Sphingopyxis alaskensis	*acdABH* – multiple copies*; acdL; etfBA; acdQP;* Sala_0761-60	Sala_0379*;* Sala_0297*;* Sala_1639-42*;* Sala_1536
α: Rhodobacterales (9 genomes)	acdABH; acdL (4); etfBA (1); hbdA (1)	

TABLE 3. Reconstructed FA degradation regulons in γ- and β-proteobacteria.

Taxonomic group[1] / organism	PsrA / FadR / FadP [2]	FA degradation genes[3]	Other co-regulated genes[3]
γ: Enterobacteriales (5 genomes)	R	*fadBA; fadH; fadD; fadIJ; fadE; fadL* (4)	*fabB* (4); *fabA* (3); *ybaW* (4); *iclR* (2); *aceBA* (1)
γ: Pasteurellales (7 genomes)	R	*fadL*	*fabA*
~Haemophilus influenzae	R	-	*fabA*
~Haemophilus ducreyi	R	*fadL*	*fabA, fabB, fabDG, accA*
~Actinobacillus pleuropneumoniae	R	*fadL*	*fabA, fabB, fabI, accAD*
γ: Alteromonadales			
Shewanella spp. (13 genomes)	A,R	*fadBA; fadH; fadD; fadIJ; fadE; acdH; etfBA* (10); *etfD* (11); *echH*; *fadL*	*aceBA; sdh;* SO2935 (11); SO0881 (12); SO4716 (11); SO0572; SO0080 (6)
Idiomarina loihiensis	A,R	*fadBA; fadH; fadIJ; fadE; acdH*	SO2935; IL0656
Colwellia psychrerythraea	A,R	*fadBA; fadD; fadIJ; fadE; acdH; etfBA/D*	SO2935
Pseudoalteromonas haloplanktis	R	*fadIJ; fadH*	
Pseudoalteromonas atlantica	R	*fadBA; fadIJ*	
Saccharophagus degradans	A	*fadBA; acdH*	
γ: Vibrionales			
Vibrio cholerae	A,R	*fadBA; fadH; fadD; fadIJ; fadE1; fadE2*	*mdh; plsB*
Vibrio parahaemolyticus	A,R	*fadBA; fadH; fadD; fadIJ; fadE1; fadE2; fadL*	*mdh; plsB*
Vibrio vulnificus	A,R	*fadBA; fadH; fadD; fadIJ; fadE1; fadE2*	*mdh; plsB*
Photobacterium profundum	A,R	*fadBA; fadH; fadD; fadIJ; fadE1; fadE2; fadL1; fadL2*	*echH*
Vibrio fischeri	R	*fadBA; fadH; fadE; fadL*	*plsB*
γ: Pseudomonadales (5 genomes)	A	*fadBA; fadH* (2); *acdH; etfBA/D; scp*	*rpoS; algQ;* PA1831; PP4195 (4)
γ: Xanthomonadales (3 genomes)	A	*acdAB*	*fabAB*; *accBC*; *aroQ*
γ: Oceanospirillales			
Hahella chejuensis	A	*fadBA; acdH; etfBA/D; scp*	
Chromohalobacter salexigens	A	*fadBA; fadH; fadD; fadE; acdH; etfBA/D; scp*	
Alcanivorax borkumensis	A	*fadD; acdH; etfBA/D*	
β: Chromobacterium (1 genome)	A	*acdAB; fadD; fadE; acdH; etfBA/D; fadL*	*mdh; accBC; fabK; fabF; aroQ*
β: Rhodocyclales			
Dechloromonas aromatica	A	*fadBA; acdAB; fadD; fadE; acdH; fadL*	
Azoarcus sp.	A	*fadBA; acdAB; fadE; fadL; echH*	
β: Burkholderiales			
Bordetella spp. (3 genomes)	A,P	*fadBA; acdAB; fadD; fadE; fadL*	**BP3678**
Ralstonia solanacearum	A,P	*fadE;* ***fadAB; acdABH; fadD; etfAB; etfD; echIH***	***acdQP; maoC; bktB; alkK***
Ralstonia eutropha, R.metallidurans	A,P	***fadAB; acdABH; fadD; etfAB; etfD; echIH, acsA***	***acdQP; maoC; bktB; alkK;*** *fab*
Burkholderia xenovorans	A,P	***fadAB; acdABH; fadD1; etfAB; etfD; echIH;*** *fadA1; fadD2; fadE*	***acdQP; maoC;*** *fab*
Burkholderia spp. (4 genomes)	A,P	***fadAB; acdABH; fadD1; etfAB; etfD; echIH;*** *fadD2*	***acdQP; maoC;*** *fab*
Methylibium petroleiphilum	P	***fadAB; acdABH; etfAB; etfD; echH***	
Polaromonas sp.	P	***fadAB; acdABH; etfAB***	
Rhodoferax ferrireducens	P	***fadAB; acdABH; etfAB; etfD***	***acdQP***

TABLE FOOTNOTES

TABLE 1.
[1]Number of complete genomes with the analyzed regulon(s) is shown in parenthesis.
[2]Existence and copy number of candidate regulators of ILV degradation genes, LiuR, and/or LiuQ, is shown by "R" and/or 'Q', respectively.
[3]Candidate LiuR-regulated genes are shown in gray background. Candidate LiuQ-regulated genes are underlined. Genes are grouped (form stretches) only when they have the same first three letters in their names. Operon structure of genes is not shown. For candidate regulon members that are not always regulated within a taxonomic group, the number of genomes that have regulated orthologs is given in parenthesis.
Functional roles of genes are listed in Table S7 in supplementary materials. The genome-specific locus tags of candidate regulon members are given in Supplementary Tables S1-S2.

Table 2.
[1]Number of complete genomes with the analyzed regulon(s) is shown in parenthesis.
[2]Candidate LiuR$^{\alpha}$-regulated genes are shown. Operon structure of genes is not shown. For candidate regulon members that are not always regulated within a taxonomic group, the number of genomes that have regulated orthologs is given in parenthesis.
Functional roles of genes are listed in Table S8 in supplementary materials. Locus tags are given for groups of orthologous genes that lack a conventional common name. The genome-specific locus tags of candidate regulon members are given in Supplementary Table S3.

Table 3.

[1]Number of complete genomes with the analyzed regulon(s) is shown in parenthesis.
[2]Existence and copy number of candidate regulators of FA degradation genes, FadR, PsrA, and FadP is shown by 'R', 'A', and 'P', respectively.
[3]Candidate FadR-regulated genes are underlined. Candidate PsrA-regulated genes are shown in gray background. Candidate FadP-regulated genes are shown in bold. Operon structure of genes is not shown. For candidate regulon members that are not always regulated within a taxonomic group, the number of genomes that have regulated orthologs is given in parenthesis.
Functional roles of genes are listed in Table S8 in supplementary materials. Locus tags are given for groups of orthologous genes that lack a conventional common name. The genome-specific locus tags of candidate regulon members are given in Supplementary Tables S4-S6.