Use of In Vivo Expression Technology To Identify Genes Important in Growth and Survival of *Pseudomonas fluorescens* Pf0-1 in Soil: Discovery of Expressed Sequences with Novel Genetic Organization

Mark W. Silby and Stuart B. Levy*

Center for Adaptation Genetics and Drug Resistance, Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts

Received 21 May 2004/Accepted 26 July 2004

Studies were undertaken to determine the genetic needs for the survival of *Pseudomonas fluorescens* Pf0-1, a gram-negative soil bacterium potentially important for biocontrol and bioremediation, in soil. In vivo expression technology (IVET) identified 22 genes with elevated expression in soil relative to laboratory media. Soil-induced sequences included genes with probable functions of nutrient acquisition and use, and of gene regulation. Ten sequences, lacking similarity to known genes, overlapped divergent known genes, revealing a novel genetic organization at those soil-induced loci. Mutations in three soil-induced genes led to impaired early growth in soil but had no impact on growth in laboratory media. Thus, IVET studies have identified sequences important for soil growth and have revealed a gene organization that was undetected by traditional laboratory approaches.

*Pseudomonas* spp. are common inhabitants of soil and water. In terrestrial environments, they can be found in close association with the plant root zone (rhizosphere) and living free in soil. Further, the fluorescent pseudomonads can also be found in association with plants and animals as commensals and pathogens. The observation that some species, such as *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*, can be found in a range of environments points to the evolution of an array of strategies to cope with environmental perturbations. This idea is supported by the description of a large number of known and predicted environmentally responsive regulators in the complete genome sequence of *P. aeruginosa* (43).

*Pseudomonas* spp. are being evaluated for applications in biocontrol, plant growth promotion, and bioremediation. Coupled with the phenotype of interest, the presumed adaptability of these isolates makes them ideal choices. However, relatively little is known about the factors contributing to their success in soil and in the rhizosphere. Soil is a complex and challenging environment with a variety of fluctuating conditions, such as nutrient level, temperature, water content, and pH. To survive and persist in soil, microbes must be able to respond and adapt rapidly to the environmental changes.

Our laboratory has been investigating the genetic basis of environmental persistence by using *P. fluorescens* Pf0-1. Selection for adherence defects led to the discovery of AdnA, a regulator affecting flagellum production, motility, attachment to sand and seeds (8, 9), and biofilm formation (5). A mutation in *adnA* reduced the ability of Pf0-1 to spread and persist in soil (21).

To further understand the underlying basis for Pf0-1 survival in soil, we sought to identify genes whose expression is induced in soil. The promoter-trapping strategy, termed in vivo expression technology (IVET), was chosen for the study. IVET was described in 1993 by Mahan et al. (19) and is similar to a method reported earlier by Osbourn et al. (27). Such promoter traps allow the positive selection of niche-induced promoters by using genetic complementation of a conditionally lethal mutation, mediated by the fusion of an in vivo-expressed promoter to the coding sequence that complements the mutation. The advantage of the IVET approach over other genetic methods is that it allows the selection of environmentally induced sequences independent of whether the loss of those sequences would be lethal. Thus, IVET helps recover both essential and nonessential genes that contribute to ecological success.

Rainey and Preston (34) suggested that the requirement for diaminopimelate and lysine by *dapB* mutant bacteria would be a useful phenotype around which to base an IVET selection. Diaminopimelate is essential but is not present in soils, and diaminopimelate auxotrophy is lethal to growing cells, while nongrowing cells can remain viable for long periods. The selection is therefore not too stringent to prevent the recovery of genes that are not induced immediately upon introduction to the soil. Here, we describe the development of a *dapB*-based IVET system and its application to unraveling the genetic basis of ecological success of *P. fluorescens* Pf0-1 in soil.

**MATERIALS AND METHODS**

**Bacterial strains, plasmids, and culture conditions.** The bacterial strains and plasmids used in this study are listed in Table 1. *P. fluorescens* and *Escherichia coli* were grown at 30 and 37°C, respectively. *E. coli* was routinely cultured in Luria-Bertani medium (39), while *P. fluorescens* strains were grown in either LB medium or *Pseudomonas* minimal medium (PMM) (15). When required, the media were solidified by the addition of 1.5% agar. Antibiotics and supplements were added as required, at the following concentrations (unless otherwise stated): ampicillin, 50 μg/ml; kanamycin, 25 μg/ml; streptomycin, 25 μg/ml; tetra-
cyclohexane, 15 μg/ml for plasmids in E. coli and 5 μg/ml for IVET plasmids integrated into the P0I-1 genome; and X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside), 50 μg/ml. Diaminopimelic acid (DAP) was added at 1 mg/ml, and lysine was used at 100 μg/ml. E. coli strains were transformed by electroporation with a Bio-Rad Micropulser following the recommendations of the manufacturer, while DNA was introduced to P. fluorescens strains by conjugation from E. coli S17-1.

DNA manipulation and sequencing. Recombinant DNA technologies were carried out as described previously (39). Restriction and DNA-modifying enzymes were purchased from Invitrogen, Inc. (Carlsbad, Calif.) and Promega (Madison, Wis.). Recombinant plasmids were routinely harvested from E. coli DH5α aprI. Plasmid DNA was prepared by using a QIAprep Spin Miniprep Kit (Qiagen, Valencia, Calif.), and total genomic DNA was prepared by using a Wizard genomic DNA purification kit (Promega). When required, DNA was recovered from agarose gel slices by using a QIAEX II gel extraction kit (QIAGEN). PCR reactions were carried out by using the Triple Master system (Eppendorf AG, Hamburg, Germany) and total genomic DNA was prepared by using a Wizard genomic DNA purification kit (Promega). When required, PCR products were cloned with pGEM-T Easy (Promega) according to the manufacturer’s instructions. Oligonucleotides were synthesized, and DNA sequences were determined by the dideoxy chain termination method (40) at the Tufts University Core Facility.

Construction and characterization of a dapB deletion in P0I-1. The region of the P0I-1 genome specifying DapB was identified from the draft genome sequence (http://genome.igi-psf.org/draft_microbes/pse/PseI_home.html) by a Blast search using dapB of P. aeruginosa PA01 as the query sequence. As in other fluorescent pseudomonads, the locus had probable dnap3, dapB, and cutx4 genes. The dapB gene was deleted by amplifying sequences flanking the open reading frame (ORF) and ligating them together, thus removing dapB. PCR primer pair DapD1 (5′-GGAGAACGATCTTGAAGGAG-3′) and DapD2 (5′-GGGAATCCATCGCCGAACGGCTTCCAG-3′) was designed to amplify 991 bp upstream (separated from the dapB ORF by 15 bp), while DapD3 (5′-GGGATCCGATGTCTTCAGAAGCCGACCCTC-3′) and DapD4 (5′-GTCTTGCTGCGCCGAGAACGACTCGT-3′) were specific for 924 bp downstream of dapB. The reverse primer for the upstream sequence and the forward primer for the downstream sequence both had a BamHI site included at the 5′ end (underlined). Amplicons were digested with BamHI, ligated together, and used as the template in a second PCR with primers DapD1 and DapD4, resulting in the in vitro generation of a dapB deletion. The deletion amplicon was cloned in pGEM-T Easy, and then cloned in the NotI site of the suicide plasmid pSR7a by utilizing the NotI sites flanking the cloning site in pGEM-T Easy. The resulting plasmid (pSRdap) was transferred to P01-1 by conjugation, and the dapB region in P01-1 was replaced with the deletion by allelic exchange (see below). Sucrose-tolerant recombinant colonies (i.e., those that had lost the suicide plasmid) were tested for DAP auxotrophy by replica plating onto PMM with and without DAP and lysine supplements. DAP auxotrophy confirmed the successful deletion of dapB. PCR with DapD1 and DapD4 amplified a product that was consistent in size with the deletion of dapB and could be cleaved by BamHI, resulting in DNA fragments of the same size as the upstream and downstream PCR products. The absence of dapB was further confirmed by the inability to amplify a sequence within the dapB ORF from the deletion strain, which was amplifiable from the parental strain (data not shown).

The deletion of dapB rendered P01-1 dependent on exogenous DAP and lysine. For a strain to be of use in an IVET screening, gradual death, or at least an inability to grow in the environment of interest, is essential. P01-1 dapB was unable to proliferate in soil, and the population slowly declined over 7 days. Over the same time period, the wild-type P01-1 population increased by 4 orders of magnitude (data not shown).

Construction of an IVET vector and library. The promoter-trapping plasmid was constructed by inserting a promoterless dapB gene into the universal IVET construct pUIC3 (32). pUIC3 has a number of features that make it ideal for construction of an IVET reporter plasmid. First, it has promoterless lacZ genes with restriction sites located 5′ of the lacZ Shine-Dalgalno sequence to allow cloning of the promoterless reporter genes. Second, a unique BglII recognition site is located upstream of the sites used for cloning reporter genes, allowing the cloning of random genomic DNA fragments upstream of the reporter genes. Finally, pUIC3 is dependent on the Pir protein for replication, so it is not maintained as an autonomous replicon in strains that do not produce Pir. We used primers DapSPF (5′-ggACTAGTCTTCTTGCCCGAGTCTGTAAGG-3′) and DapSPR (5′-ggACTAGTCTTCAAGCCGGACGACGTC-3′) to amplify the dapB ORF, along with the probable Shine-Dalgalno sequence at the 5′ end, from P01-1. SpeI restriction sites (underlined) incorporated into the 5′ end of the primers allowed cloning of the dapB′ amplicon into the unique SpeI site of pUIC3, upstream of the promoterless lacZ genes. Thus, pIVETdap has a promoterless dapB-lacZ transcriptional fusion.

Four fusion libraries were constructed with pIVETdap. The first of these was constructed by cloning BglII-digested total DNA into the corresponding site of pIVETdap. The other three libraries were constructed by isolating 1.5- to 2.5-kb fragments of P01-1 DNA, partially digested with Sau3AI, from an agarose gel and ligating these into the BglII site of pIVETdap. After the transformation of E. coli S17-1 with the libraries, 20 random colonies were chosen and the plasmid DNA was prepared. Restriction digestion indicated that 90% of the isolates had inserts, and a range of insert sizes had been cloned, supporting the description as a random genomic library (data not shown).

Selection of soil-induced promoters. Four genomic libraries were constructed in pIVETdap, and each was transferred to P01-1 dapB by conjugation. Since pIVETdap cannot replicate in P01-1, we concluded that colonies that had recovered had library clones integrated into the P01-1 genome by recombination.
between the plasmid-borne and chromosomal copies of the cloned fragment. Pooled transconjugant colonies arising from each library were used to inoculate two separate soil samples (see below for soil procedures). Each inoculum consisted of cells from approximately 1,500 transconjugant colonies. After 7 days of selection, the transconjugants were recovered and plated onto PMM supplemented with DAP, lysine, tetracycline, and X-Gal. Bacteria that survived selection produced colonies; white ones were chosen for further use.

Soil growth and survival assay. Strains were grown for 20 h in PMM broth with appropriate antibiotics and growth supplements. The soil used was a gamma-irradiated fine sandy loam, which has been described previously (9). Bacteria were inoculated to an approximate population of 2 × 10^8 CFU/ml in sterile distilled H_2O, and 1 ml was mixed with 5 g of sterile soil (achieving approximately 50% water-holding capacity). After 30 min, a 1-g sample was removed to allow the initial recoverable CFU to be determined. Subsequent samples were taken when required. CFU of soil was determined by adding 1 ml of sterile H_2O to 1 g of soil and vortexing vigorously for 1 min. The soil suspension was allowed to settle briefly so that any large particles fell out of suspension. The soil suspension was then serially diluted 10-fold and plated on selective media to allow enumeration of viable, cultivable bacteria by colony counting on PMM plates with appropriate supplements to support growth.

Construction of plasmids for knockout of iiv2, iiv3, and iiv7. A wild-type DNA fragment was amplified from each locus by using Pbo-1 total DNA as a template, with primers iiv2F (5′-CCGAGCAACTGATCTCCCA) and iiv2R (5′-CGGCGATATTTTCGATAGGT); iiv3F (5′-GACGATCGGGTGTTCACTTT) and iiv3R (5′-AGGTCCTGCGGATGGA); and iiv7F (5′-ATGAAATCTCTTTGGTTCG) and iiv7R (5′-TCAGCACTGCGTCTCGG). Each product was then cloned with pGEM-T Easy. Gene disruptions were generated by making partial deletions and inserting a Sm^r omega cassette as follows. The iiv2 clone (pGEMiiv2) was digested with Msel-cut and dephosphorylated with calf intestinal alkaline phosphatase [CIAP]), for which there are recognition sites at positions 2997 and 3976 of the 5,535-bp iiv2 ORF. A Sm^r cassette was liberated from pHRP317 with EcoRV and Smal, gel purified, and ligated to the Msel-cut pGEMiiv2, generating the plasmid pGEMiiv2Sm in which a Sm^r cassette replaces 973 bp of iiv2 sequence between two Msel sites. The iiv3 clone (pGEMiiv3) was digested with BamHI, which cuts the 1,560-bp iiv3 ORF sequence at base pair positions 744 and 1176, and treated with CIAP. A Sm^r cassette with BamHI ends was cut from pHPR315, gel purified, and ligated to the BamHI-digested pGEMiiv3, resulting in pGEMiiv3Sm. This plasmid has a 426-bp deletion between two BamHI sites in the iiv3 coding sequence, which is replaced by the Sm^r cassette. The sequence amplified in the iiv7 clone (pGEMiiv7) contained the ORFs for the iiv7^+ response regulator and sensor kinase (see Results). It was digested with ClaI (by using DNA prepared from the dam-negative strain GM2163), end-filled with Klenow fragment, and treated with CIAP. ClaI cuts the cloned sequence at 398 and 1,095 bp. A Sm^r cassette cut from pHPR317 with EcoRV and Smal was ligated to the free ends of the digested pGEMiiv, creating pGEMiivSm. This plasmid lacks the last 287 bp of iiv7 sequence regulator gene and the first 340 bp of the 1,416-bp sensor kinase gene and has a Sm^r cassette in place of the deleted sequence. Finally, each of the iiv3:Sm sequences in pGEM-T Easy clones was released by digestion with NotI and ligated with NotI-digested pSR47s (23), resulting in the suicide constructs pSRiiv2Sm, pSRiiv3Sm, and pSRiive7Sm.

Allele exchange mutagenesis. Mutated alleles cloned into the suicide plasmid pSR47s (see above) were transferred to P. fluorescens Pbo-1 by conjugation, by using E. coli ST174 as the donor strain. Transconjugants resistant to antibiotic markers carried by the plasmid (Km^r) and, in the case of iiv mutations, to the mutated allele (Sm^r), were selected. Since pSR47s cannot replicate in P. fluorescens, the growth of transconjugants indicated that the entire vector carrying the disrupted allele had integrated into the chromosome by recombination. The recombinant bacteria were grown overnight in LB broth containing ampicillin and plated on LB agar supplemented with ampicillin, streptomycin (for Sm^r knockouts), and 5% sucrose. Growth of iiv knockout candidates indicated that the mutated allele was present (Sm^r) and that the plasmid vector had been lost, the growth of transconjugants indicated that the entire vector carrying the mutated allele was present (Sm^r) and that the plasmid vector had been lost.

RESULTS AND DISCUSSION

Selection and sequencing of soil-induced promoters. Growth of the Pbo-1 IVET library-bearing strains in soil (see Materials and Methods) led to the recovery of active promoter fusions to dapB. Prior to soil inoculation, 31% of transconjugants were blue on X-Gal. However, greater than 99% of recovered bacteria gave rise to blue colonies, indicating that growth in soil had indeed enriched the population for active fusions, leading to the expression of lacZY. Among the recovered colonies, those that were white or very pale blue (<1%) possessed putative soil-induced promoters since they indicated low expression of lacZY in laboratory medium. These candidates were tested for dependence upon DAP and lysine. In total, 26 strains that survived 7 days in soil but had DAP and lysine dependency in laboratory medium were isolated. Each strain was reintroduced separately into soil to confirm its ability to proliferate compared to the parental Pbo-1 ΔdapB strain (Fig. 1). Each strain was able to grow in soil significantly better than the ΔdapB mutant, which was consistently unable to multiply, supporting the suggestion that these isolates have soil-induced promoters fused to the promoterless dapB. However, none of these strains were fully complemented to wild-type levels by the promoter fused to dapB, which may indicate that this screening was useful in recovering promoters with moderate (less than native dapB) expression levels.

Plasmids harboring IVET fusions were recovered from their hosts by using conjugation, as described previously (33). The primer IVET3 (5′-GCCTCGACAGAATCTTCG) was designed to anneal near the start of dapB reading toward the start codon and was used to determine the DNA sequences fused to dapB. Among the 26 fusions, we identified 22 different sequences; four sequences (iiv7, iiv9, iiv14, and iiv20) were isolated twice. Although originally from the same library, the duplicate fusion sequences were isolated from separate soil selection experiments, supporting their being genuine soil-induced fusions. The 22 different sequences obtained fell into two distinct groups. The first group, which we call positive fusions, has similarity to known or predicted genes in the correct orientation to drive the expression of dapB (Table 2). The second group showed similarity to known or predicted genes on the DNA strand in the opposite direction of dapB (Table 3). It has been suggested that the sequences in the second group might harbor previously unrecognized promoters fused to dapB, and therefore these have been called cryptic fusions (41).

Analysis of DNA sequences fused to dapB in positive IVET fusions. (i) Genes related to nutrition. Possible roles in the acquisition and utilization of nutrients are revealed by six positive fusions (iiv3, iiv11, iiv6, iiv15, iiv18, and iiv20; see Table 2), suggesting that the soil used has a number of useful compounds that can support growth but that these are not found in standard laboratory media. We suspect that many biosynthetic genes would also be expressed in soil but would have been excluded from this study because of the use of a minimal medium for selection. The predicted product of iiv3 is similar
to FadD acid-coenzyme A (CoA) ligases, and has 35% identity (51% positive) with FadD of \textit{E. coli} (GenBank accession number L02649). Cloning and expression of \textit{E. coli} fadD demonstrated that the product was responsible for the acyl-CoA synthetase activity, which esterifies fatty acids into metabolically active CoA thioesters (2). This reaction makes the fatty acid available for β-oxidation and thus a useable carbon and energy source. In \textit{iiv11}, the Pf0-1 subtype 302 gene, predicted to encode a xylanase/chitin deacetylase, is found. Interestingly, upstream of the gene is another predicted xylanase/chitin deacetylase gene (Pflu5303), the 3′ end of which overlaps the 5′ end of Pflu5302 by 8 bp. It is likely that these two xylanase/chitin deacetylase genes are transcribed together, possibly from a promoter upstream of Pflu5303. The induction of \textit{iiv11} may reflect the presence of xylans and/or chitin, of plant or insect origin, respectively, whose degradation would yield a useful carbon source.

Four potential transporters were upregulated during growth in soil (\textit{iiv6}, \textit{iiv15}, \textit{iiv18}, and \textit{iiv20}). The \textit{iiv6} sequence is predicted to specify a Na+/serine symporter. Unlike the other positive fusions, the sequence juxtaposed to \textit{dapB} in \textit{iiv6} was not part of a gene annotated in the Pf0-1 draft genome sequence. Rather, the DNA cloned in the fusion ended 194 bp upstream of the predicted Na+/serine symporter (Pflu4867). An examination of the 1,486 bp of intergenic sequence upstream of the Pflu4867 gene with SoftBerry software suggests the presence of two predicted promoters, both in the correct orientation to drive expression of Pflu4867 and both within the...
region contained in the *iiv6* fusion. The predicted transcriptional start site from the promoter nearest to the symporter ORF starts 259 bp upstream of the ORF.

*iiv15* is predicted to specify a cytosine/purine/uracil/thiamine/allantoin permease (Pflu4595). An NCBI conserved-domain search using the position-specific scoring matrix (PSSM) indicates that Pflu4595 aligns with CodB (7). However, Pflu4595 appears to be about 90 amino acids shorter than the other members of that group.

A gene that could encode a protein with export or import function is found in *iiv18*, a member of the clustered orthologous group 3451 (COG3451) type IV secretory pathway, VirB4 components group. NCBI conserved-domain searching indicates that there is some significant alignment with a VirB4 consensus, but that this alignment is only between amino acid residues 326 and 799, which represents just 52.3% of the protein. Using Blastp, we found that the nearest match in public databases is with a conserved hypothetical protein with a VirB4 domain (CAE92911). Blastp searches with the amino acid sequences before and after the VirB4 consensus failed to give further insight into potential functions. VirB4 is an ATPase, and VirB4-like proteins are part of the plasmid conjugation machinery and the type IV secretion apparatus. In addition, competence (DNA import) in *Helicobacter pylori* requires homologues of a number of conjugation proteins, including VirB4.

It is interesting that the genes found in fusions *iiv15* and *iiv18* may both, conceivably, function in the transport of nucleotides or DNA. The increased expression of *codB* in nitrogen-limited conditions (25) suggests that import of nitrogen-containing compounds, such as cytosine, purines, uracil, thiamine, and/or allantoin, may supply a nitrogen source. Although generally considered with respect to acquiring new genetic information, importing DNA also provides a good nitrogen and carbon source, e.g., for *E. coli* growing in starvation cultures (10).

The *iiv20* sequence is predicted to specify a member of the major facilitator superfamily of permeases. The *iiv20* sequence has similarity with the sequence encoded by PcaT, which is thought to function as a system by which starved cells can scavenge β-ketoadipate from the environment (31). In Pf0-1

### TABLE 2. Analysis of positive fusions

<table>
<thead>
<tr>
<th>Fusion</th>
<th>Pflu&lt;sup&gt;a&lt;/sup&gt;</th>
<th>COG</th>
<th>Best hit&lt;sup&gt;b&lt;/sup&gt;</th>
<th>PA01 match (% identity, % positive substitutions)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
</table>
| Nutrient utilization  
*iiv3* | Pflu5341 | COG0318, acyl-CoA synthetases (AMP-forming)/AMP-acid ligases | Putative ligase; *Bordetella parapertussis*; E = e<sup>-138</sup> | PA3300 (29, 45) |
|  
*iiv11* | Pflu5302 | COG0726, predicted xylanase/chitin deacetylase | Putative saccharide deacetylase; *Sinorhizobium meliloti*; E = e<sup>-29</sup> | PA1517 (26, 40) |
| Transport  
*iiv6<sup>d</sup>* | Pflu4867 | COG3633, Na<sup>+</sup>/serine symporter | Sodium-dicarboxylate symporter family; *P. putida* KT2440; E = 0.0 | PA2042 (80, 87) |
|  
*iiv15* | Pflu4596 | COG1457, purine-cytosine permease and related proteins | Cytosine/purine/uracil/thiamine/allantoin permease; *P. putida* KT2440; E = e<sup>-174</sup> | PA5099 (35, 53) |
|  
*iiv18* | Pflu5167 | COG3451, type IV secretory pathway, VirB4 components | Conserved hypothetical protein with VirB4 domain; *P. putida*; E = 0.0 | PA0229 (75, 84) |
|  
*iiv20<sup>c</sup>* | Pflu0451 | COG0477, permeases of the major facilitator superfamily | Permeases of the major facilitator superfamily; *P. syringae pv. syringae*; B728a; E = 0.0 | PA0229 (75, 84) |
| Regulation  
*iiv7<sup>d</sup>* | Pflu4942 | COG0642, signal transduction histidine kinase | Signal transduction histidine kinase; *Burkholderia fungorum*; E = e<sup>-121</sup> | PA4886 (45, 62) |
|  
*iiv9<sup>d</sup>* | Pflu2333 | COG1167, MocR family transcriptional regulators | COG1167: transcriptional regulator; *P. syringae pv. syringae* B728a; E = 0.0 | PA1654 (80, 88) |
| Detoxification and metabolism  
*iiv17* | Pflu4220 | COG0604, NADPH quinone reductase and related Zn-dependent oxido-reductases | Putative oxidoreductase; *Streptomyces avermitilis* MA-4680; E = e<sup>-108</sup> | PA3567 (31, 48) |
|  
*iiv21* | Pflu1840 | COG0315, molybdenum cofactor biosynthesis enzyme | Molybdopterin biosynthetic protein C; *P. aeruginosa*; E = 5e<sup>-60</sup> | PA3918 (75, 86) |
| Hypothetical  
*iiv2* | Pflu5100 | None | None | None |
|  
*iiv24<sup>e</sup>* | None | None | Hypothetical protein; *P. putida* KT2440; E = 5e<sup>-19</sup> | None |

---

<sup>a</sup> This fusion was isolated twice in the IVET screen for soil-induced promoters.

<sup>b</sup> The fusion is likely to contain the promoter for Pflu4867. None of this ORF is present in the fusion.

<sup>c</sup> Pflu numbers are the NCBI designations for genes identified in the NCBI annotation of the Pf0-1 genome.

<sup>d</sup> Database searches were carried out by using Blasts with the Pflu sequences as queries. Best hit excluding matches to Pf0-1 is shown.

<sup>e</sup> Closest match in the *P. aeruginosa* PA01 genome sequence (http://www.pseudomonas.com).
and Pseudomonas putida, pcaT is found among other genes specifying factors of importance in the protocatechuate branch of the \(\beta\)-ketoadipate pathway. Therefore, the trapping of pcaT could relate to the induction of an upstream gene that is important for another part of the \(\beta\)-ketoadipate pathway.

While the induction of genes involved in nutrient acquisition and utilization solves a basic need for nutrient sources, an additional role in environmental sensing cannot be excluded. O’Toole et al. proposed that the catabolite repression control protein (Crc) in *P. aeruginosa* integrates nutritional cues as part of a pathway to regulate biofilm development (28). In *P. fluorescens* WCS365, biofilm formation proceeds via a number of convergent pathways responding to environmental (nutrient) factors (29). In *Pseudomonas aureofaciens*, low concentrations of inorganic phosphate lead to the expression of the Pho regulon, which is involved in scavenging phosphate from the environment and repression of biofilm formation (24).

(ii) Regulatory genes (iv7 and iv9). In iv7, the sequence belongs in the signal transduction histidine kinase (COG0064) family. The predicted translation product shows similarity to heavy metal sensor histidine kinases, such as CzcS (which controls cobalt, zinc, and cadmium resistance) of *Ralstonia solanacearum* (gi 17430958). In support of this prediction, the sequence upstream (Pflu4943) is similar to CzcR, the response regulator partner for CzcS. Despite the possible role in metal resistance, efforts to induce expression of iv7 in laboratory media by exposure to sublethal concentrations of copper, zinc, and cobalt were unsuccessful (data not shown). This failure may indicate a specific inducer in the soil environment.

The gene identified in iv9 is predicted to encode a member of the GntR family of regulators, which contain a DNA-binding helix-turn-helix domain and an aminotransferase domain. This family of regulators was first described by Haydon and Guest (13) and has been further refined into subfamilies (37). The sequence found in iv9 most closely fits with the MocR subfamily defined by Rigali et al. (37).

(iii) Genes reflecting detoxification and metabolism. Although we hypothesize that soil represents a harsh environment and repression of biofilm formation on survival.

The predicted product for the gene identified at the iv21 locus shows strong similarity with predicted molybdenum cofactor biosynthesis protein C (MoaC) from *P. aeruginosa* (PA3918; 75% identity) and *Pseudomonas syringae* (accession number ZP_00125576; 72% identity). However, the similarity is apparent only between residues 98 and 254. iv21 appears to specify a protein that is 93 amino acids longer than many other members of the COG0315 group. An exception is a predicted protein from *Azotobacter vinelandii* (accession number ZP_00089475), which is of a similar size as the iv21 gene. In *E. coli*, MoaC is involved in the synthesis of a molybdopterin precursor, which is ultimately converted to molybdopterin guanine dinucleotide (35). The molybdenum cofactor can be found in at least nine enterobacterial anaerobic enzymes (12).

(iv) Hypothetical genes. The iv2 sequence does not match any sequence in GenBank; the 1,844-amino acid predicted translation product has no significant (E < \(10^{-10}\)) matches to sequences in public databases, and no conserved domains are evident (NCBI conserved domain search). The iv24 sequence is annotated as hypothetical by the Joint Genome Institute but does not appear in the NCBI annotation of the Pf0-1 genome. The predicted translation product (113 amino acids) has 61% identity with another hypothetical protein from *P. putida* KT2440 between amino acids 33 and 109 but no matches that span the full-length sequence. Using the SignalP server (http://www.cbs.dtu.dk/services/SignalP-2.0/) (26), we found that the first 20 amino acids are predicted to be a signal peptide, cleaved between amino acids 20 and 21. Neither the iv2 nor iv24 predicted proteins can be matched with a COG. The novel nature of these sequences supports the contention that

### TABLE 3. Analysis of “cryptic fusions”

<table>
<thead>
<tr>
<th>Fusion</th>
<th>ORF length (bp)</th>
<th>Codon usage %</th>
<th>Predicted promoter</th>
<th>Gene on opposite strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>iv1</td>
<td>627</td>
<td>15.77</td>
<td>Yes</td>
<td>Pflu5418: COG0604: NADPH quinone reductase and related Zn-dependent oxidoreductases</td>
</tr>
<tr>
<td>iv4</td>
<td>1,569</td>
<td>16.98</td>
<td>No</td>
<td>Pflu1118: COG1902: NADH flavin oxidoreductases, Old Yellow</td>
</tr>
<tr>
<td>iv5</td>
<td>1,038</td>
<td>16.67</td>
<td>Yes</td>
<td>Pflu1358: COG1012: NAD-dependent aldehyde dehydrogenases</td>
</tr>
<tr>
<td>iv8</td>
<td>1,701</td>
<td>16.39</td>
<td>No</td>
<td>Pflu2278: COG0855: polyphosphate kinase</td>
</tr>
<tr>
<td>iv12</td>
<td>1,308</td>
<td>17.7</td>
<td>Yes</td>
<td>Pflu2679: COG2081: predicted flavoprotein</td>
</tr>
<tr>
<td>iv13</td>
<td>213</td>
<td>27.47</td>
<td>No</td>
<td>Pflu5361: COG0583: transcriptional regulator (LysR type)</td>
</tr>
<tr>
<td>iv14</td>
<td>1,020</td>
<td>14.81</td>
<td>Yes</td>
<td>Pflu5906: COG1289: predicted membrane protein. Similar to predicted fusaric acid resistance proteins</td>
</tr>
<tr>
<td>iv19</td>
<td>1,908</td>
<td>17.7</td>
<td>Yes</td>
<td>Pflu5051: COG0119: isopropylmalate/homocitrate/citramalate synthases</td>
</tr>
<tr>
<td>iv23</td>
<td>321</td>
<td>20.78</td>
<td>No</td>
<td>Pflu3174: COG0584: glycophosphoryl diester phosphodiesterase</td>
</tr>
<tr>
<td>iv25</td>
<td>267</td>
<td>23.39</td>
<td>No</td>
<td>Pflu3727: COG0189: glutathione synthase/ribosomal protein S6 modification enzyme (glutaminyl transferase)</td>
</tr>
</tbody>
</table>

---

*a This fusion was isolated twice in the IVET screening for soil-induced promoters.

*b Length of the cryptic gene ORF identified from IVET fusions.

c Mean difference when compared to the *P. fluorescens* codon usage table, http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species = Pseudomonas + fluorescens + [gbct].

d Putative promoters upstream of the iv ORF predicted by using SoftBerry software.

e Gene found on the opposite DNA strand relative to the cryptic genes, based upon NCBI annotation of *P. fluorescens* Pf0-1 genome sequence (NZ_AAAT0000000).
Cryptic IVET fusions. The second major class of fusions (10 out of 22) isolated in this study were those in which the sequence trapped in the same orientation as \( d\alpha pB' \) had no similarity to known genes or proteins (Blastn and Blastx searches). This group of trapped sequences had considerable overlap with previously described or predicted genes on the opposite DNA strand (Table 3 and Fig. 2). We have termed these cryptic ORFs. Thus, our analysis of the codon usage of RecA, DapB, GlnA, and FlgB from Pf0-1 revealed mean differences of 14.33, 11.84, 14.69, and 16.8%, respectively (41). Thus, the seemingly high mean differences might be attributed to variations between Pf0-1 and other \( P. \) fluorescens strains.

Interestingly, the sequence trapped in the cryptic fusion \( iiv4 \) is antisense to a predicted \( fadH \) gene (encoding 2,4-diencyl-CoA reductase) on the opposite strand, which has been shown to be required for the \( \beta \)-oxidation of unsaturated fatty acids with double bonds extending from even-numbered carbon atoms (46). The induction of such a sequence raises the possibility of RNA-mediated regulation of fatty acid utilization in soils.

Our identification of cryptic genes could indicate that the \( P. \) fluorescens genome, and perhaps other bacterial genomes, have more genes than currently thought, which would naturally imply a greater coding density in the genome sequence. Possible functions include the transcript acting as an antisense regulator of the opposite gene. In prokaryotes, antisense regulation is involved in controlling plasmid replication and bacteriophage gene expression (for a review, see reference 44), and a growing number of regulatory RNA molecules are being identified in prokaryotes (for examples, see references 14 and 22). Alternatively, the activation of the cryptic gene could shut down transcription of the opposite gene due to competition for the template. The extent to which this largely stochastic situation might be effective could depend upon the relative promoter strengths in different environments, and thus environmental fluctuations would indirectly control relative transcription at such loci.

We are not the first to report the recovery of such fusions from IVET screenings. Indeed, the first report on IVET (19) mentions a fusion in the \( rfb \) operon which when transcribed would be antisense to \( rfb \) gene transcripts; and other examples exist (4, 32, 45). However, we are the first to emphasize the significance of these fusions and to use an existing whole genome sequence to examine them further.

Role for \( iiv \) genes in proliferation in soil. The importance of proteins specified by \( iiv \) sequences for growth and/or survival in soil was tested by constructing mutants in which \( iiv \) sequences representing different classes of genes were partially deleted. We disrupted the genes found in \( iiv1 \), \( iiv3 \) (\( FadD \)), and \( iiv7 \) (two-component sensor), replacing part of each ORF with a streptomycin resistance gene cassette (see Materials and Methods for the construction). The ability of the recombinants to grow in soil was assessed by using our standard soil growth assay.

Each \( iiv::Sm \) mutant showed a significant defect in soil growth after 1 day (\( P < 0.05 \)) relative to Pf0-2x (Fig. 3). Pf0-2x is a Pf0-1 derivative harboring \( Sm^+ \) in the \( adnA \) gene (38), and was used to rule out reduced fitness linked to the \( Sm^+ \) cassette. AdnA is important for long-term persistence and is spread in natural soils under field conditions (21), but under conditions similar to those employed here, the lack of AdnA in Pf0-2x did not influence growth (9). The population of the \( iiv \) mutants increased less than 100-fold, compared to Pf0-2x, which increased over 1,000-fold after 1 day. Interestingly, after 3 days in soil, the three \( iiv::Sm \) mutants had established populations similar to Pf0-1 and Pf0-2x. Mutations in these three \( iiv \) genes clearly have an impact on soil fitness, presumably impairing early establishment of the population. This finding was not associated with a discernible growth defect in cultures of \( iiv::Sm \) mutants in vitro. Although the growth characteristics...
shown in Fig. 3 resemble those in Fig. 1A, the slow growth of fusion strains (Fig. 1) correlates with slow growth in laboratory media (data not shown) and is due to the fact that growth of fusion strains depends on the expression of dapB by heterologous promoters. In contrast, dapB in the iv mutants is driven by its native promoter, and the growth defect is a consequence of mutations in iv genes.

Most IVET-type screens to date have been focused on elucidating the mechanisms of bacterial pathogenesis in mammals (for examples, see references 4, 16, 20, and 45). More recently, IVET approaches have been used to examine the processes of plant pathogenesis (3), colonization of a plant pathogenic fungus by Pseudomonas putida (17), and rhizosphere colonization by Pseudomonas spp. (11, 32, 36). This IVET study has focused on genes induced during growth in soil. Of interest, we have found some overlap between soil-induced and rhizosphere-induced genes. Rainey (32) isolated a fusion (rhi-3) containing sequences similar to copR and copS, which are involved in resistance to copper. Upstream of iv7, a putative sensor kinase from a two-component regulatory pair, is a sequence predicted to encode a protein that is 61% identical to CopR (GenBank accession number L05176) and has 63% identity with CzcR (GenBank accession number Q44006), a regulator of resistance to cobalt, zinc, and cadmium. iv7 is likely to specify the functional equivalent of rhi-3 and CopS or CzcS. A second overlap appears between iv20 and the cii-10 fusion reported by Rediers et al. (36). In cii-10, genes with a high degree of similarity to pcaB and pcaC were found. The products of these genes are important in the protocatechuate branch of the β-ketoadipate pathway for the degradation of aromatic compounds, which is the same system in which PcaT (iv20) functions. In another example, Gal et al. identified a putative acyl CoA dehydrogenase (COG1960) in the fusion rhi-27 (11), whereas the sequence in fusion iv3 is two genes downstream from a COG1960-family acyl CoA dehydrogenase gene which is likely to be functionally equivalent to rhi-27. Finally, both iv15 and the rhi-73 fusion isolated by Gal et al. (11) specify predicted COG1457 nucleotide permeases. These similarities in recovered sequences between the rhizosphere studies and our soil work suggest that those shared gene fusions may be ones more generally responsive to the soil surrounding the rhizosphere than to the rhizosphere per se. In contrast, there is very little overlap between the small number of fusions recovered here and those identified by using IVET, differential display, and signature-tagged mutagenesis to examine gene expression in model pathogenesis systems (18).

This study, the third which utilized the requirement for DAP, uncovered two groups of fusions which are affected by soil contact. The first group has similarity to known or predicted genes, while the second does not. The second group comprises 10 cryptic genes and points to a novel genome organization that has so far been overlooked in genome annotation efforts.

ACKNOWLEDGMENTS

We thank Laura McMurry for reading the manuscript. This study was supported by Department of Energy grant DE-FG02-97ER62493.

ADDENDUM IN PROOF

The NCBI annotation of the P. fluorescens Pf0-1 genome sequence has been updated to reflect newly released data. With this update, Pf0 numbers have been changed. We anticipate that Pf0 numbers may change again before the genome sequence is completed. Anyone interested in the P. fluorescens sequences described herein should contact the authors for the most recent accession numbers.

REFERENCES


