JOURNAL OF BACTERIOLOGY, Mar. 1999, p. 1585–1602 Vol. 181, No. 5
0021-9193/99/$04.00+0

Complete Sequence of a 184-Kilobase Catabolic Plasmid from *Sphingomonas aromaticivorans* F199†

MARGARET F. ROMINE,[1]* LISA C. STILLWELL,[1] KWONG-KWOK WONG,[1] SARAH J. THURSTON,[1] ELLEN C. SISK,[1]‡ CHRISTOPH SENSEN,[2] TERRY GAASTERLAND,[3,4]§ JIM K. FREDRICKSON,[1] AND JEFFREY D. SAFFER[1]

Pacific Northwest National Laboratory, Richland, Washington 99352[1]; National Research Council of Canada, Institute for Marine Biosciences, Halifax, Nova Scotia B3H 3Z1, Canada[2]; Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439[3]; and Department of Computer Science, University of Chicago, Chicago, Illinois 60637[4]

Received 10 July 1998/Accepted 16 November 1998

The complete 184,457-bp sequence of the aromatic catabolic plasmid, pNL1, from *Sphingomonas aromaticivorans* F199 has been determined. A total of 186 open reading frames (ORFs) are predicted to encode proteins, of which 79 are likely directly associated with catabolism or transport of aromatic compounds. Genes that encode enzymes associated with the degradation of biphenyl, naphthalene, *m*-xylene, and *p*-cresol are predicted to be distributed among 15 gene clusters. The unusual coclustering of genes associated with different pathways appears to have evolved in response to similarities in biochemical mechanisms required for the degradation of intermediates in different pathways. A putative efflux pump and several hypothetical membrane-associated proteins were identified and predicted to be involved in the transport of aromatic compounds and/or intermediates in catabolism across the cell wall. Several genes associated with integration and recombination, including two group II intron-associated maturases, were identified in the replication region, suggesting that pNL1 is able to undergo integration and excision events with the chromosome and/or other portions of the plasmid. Conjugative transfer of pNL1 to another *Sphingomonas* sp. was demonstrated, and genes associated with this function were found in two large clusters. Approximately one-third of the ORFs (59 of them) have no obvious homology to known genes.

Sphingomonas aromaticivorans F199 was isolated from sediments collected 410 m below the land surface near Allendale, S.C., in 1988 (4, 21). It was established that this bacterium possessed the novel ability to degrade a variety of aromatic compounds including toluene, all isomers of xylene, *p*-cresol, naphthalene, biphenyl, dibenzothiophene, fluorene, salicylate, and benzoate (20, 22). In recent years, there have been many reports of other *Sphingomonas* strains that are capable of degrading aromatic compounds (12, 17, 30, 36, 40, 44, 45, 62–65, 68–70, 88, 91). Studies of *Sphingomonas* strains suggest that members of this genus are well adapted for the degradation of high-molecular-weight polycyclic aromatic hydrocarbons and other aromatic contaminants. The inability to detect *Sphingomonas* biodegradative genes via hybridization with catabolic genes from phylogenetically distinct bacteria suggested that biodegradative genes from *Sphingomonas* sp. evolved independently from phylogenetically distinct bacteria such as those within the genus *Pseudomonas* (46, 47). Some *Sphingomonas* strains are further distinguished in that the genes necessary for degradation of one type of aromatic compound are distributed into multiple operons that also possess genes for the degradation of other aromatic compounds (107). This unusual gene arrangement suggests that a highly complex regulatory network is responsible for the expression of aromatic degradative pathways in some *Sphingomonas* spp.

S. aromaticivorans F199 was shown to possess two plasmids (20, 22), which are designated pNL1 (~180 kbp) and pNL2 (~480 kbp). We reported earlier that catechol *meta* ring cleavage activity, a central step in the catabolism of aromatic rings, was associated with the smaller plasmid, pNL1 (86), and we described a physical map for this plasmid. To further probe the catabolic functions and accessory genes encoded on pNL1, we undertook the complete sequencing and annotation of this plasmid. This approach has allowed a thorough genetic analysis of pNL1-associated catabolic genes, a comparison of these genes with analogous chromosomally located ones in *S. yanoikuyae* B1, and the development of hypotheses regarding functions of pNL1-encoded genes.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. *S. aromaticivorans* F199, originally isolated by our laboratory, is also maintained in the U.S. Department of Energy's Subsurface Microbial Culture Collection at Florida State University (3). *Sphingomonas* sp. strain S-88 (ATCC 31554) mutant m260 was obtained from Thomas Pollock (77). Its resistance to bacitracin and inability to degrade aromatic compounds examined in this study was useful for plasmid transfer experiments. *S. paucimobilis* ATCC 298377 was purchased from the American Type Culture Collection. Bacteria were maintained on either full-strength (*Escherichia coli*) or half-strength (*Sphingomonas*) Luria-Bertani medium (LB).

SB354, a Tn*5* derivative of the suicide vector pRK600 (85) was maintained on LB containing kanamycin to select for the transposon and chloramphenicol to select for the pRK600 plasmid containing the transposon. Plasmid pRK2013 (18) was used as a helper plasmid in triparental matings. Antibiotics were supplemented at levels of 25 μg/ml for kanamycin and chloramphenicol; 500 μg/ml for bacitracin, prepared from (Sigma) powder at 73,000 U/g; and 12.5 μg/ml for polymyxin B.

Transposon mutagenesis. A portion (50 μl) of an overnight culture of the *S. aromaticivorans* F199 recipient was spotted and allowed to dry on half-strength LB agar. An equivalent amount of overnight cultures of the donor strains, *E. coli* with either SB354 or pRK2013, were then overlaid, dried, and incubated over-

* Corresponding author. Mailing address: Pacific Northwest National Laboratory, Environmental Microbiology Group, 902 Battelle Blvd., MS P7-50, Richland, WA 99352. Phone: (509) 376-8287. Fax: (509) 376-1321. E-mail: Margie.Romine@pnl.gov.

† NRCC publication 42280.

‡ Present address: Seattle Biomedical Research Institute, Seattle, WA 98109-1651.

§ Present address: The Rockefeller University, New York, NY 10021-6399.

night at 30°C. The resulting bacterial spots were resuspended in 1 ml of saline, and 100-μl aliquots were plated onto 0.5× LB containing kanamycin to select for transposition and polymyxin B to select against the *E. coli* donor strains. *Sphingomonas* colonies are readily distinguished from *E. coli* colonies because of their characteristic yellow pigmentation. The resulting *Sphingomonas* insertion mutants were screened for loss of dioxygenase activity by placing a small crystal of indole in the lid of the petri dish; the colonies were then examined for the absence of blue coloring (i.e., lack of indigo formation). Conditions for testing the ability to grow on aromatic compounds or to produce colored intermediates from them by *Sphingomonas* strains have been described elsewhere (20).

Conjugal transfer of pNL1. Recipient *Sphingomonas* sp. S-88 m260 and donor *S. aromaticivorans* F199 strains were mated as described above. Exconjugants were isolated on 0.5× LB plates containing bacitracin to select against donor cells and kanamycin to select for transfer of transposon-mutagenized pNL1. After incubation, a few crystals of indole were added to the lid of the plate and further incubated to allow the development of blue color. The presence of pNL1 in exconjugants was confirmed by plasmid isolation (39) and analysis by pulse field gel electrophoresis (1% agarose, 0.5× TBE, 6 V/cm, included angle of 120, 0.475-s initial switch time, 21.79-s final switch time, linear ramping factor, 20.5-h run time) with the Bio-Rad Chef system (Bio-Rad, Hercules, Calif.).

Construction and isolation of small and large insert libraries. The procedure for isolation of pNL1 has been described elsewhere (86). For construction of the large insert library, a partial *Sau*3A digestion of pNL1 was used to produce fragments of approximately 40 kb, which were then dephosphorylated and ligated into the *Bam*HI site of the sCos-1 cosmid vector (16). Ligation mixtures were packaged into lambda particles by using the Gigapack II packaging extract (Stratagene, La Jolla, Calif.), infected into *E. coli* DH5α, and plated onto LB agar supplemented with 100 μg of kanamycin per ml. Cosmids were isolated from 6-ml overnight cultures of the resulting kanamycin-resistant colonies by using a modified boiling lysis procedure (55). Briefly, bacterial pellets were dissolved in 300 μl of solution A (8% sucrose; 5% Triton X-100; 50 mM Tris, pH 8.0; 50 mM EDTA) and 10 μl of solution B (10 mg RNase per ml; 1 mg of lysozyme per ml; 50 mM Tris, pH 8) and boiled for 1 min. After removal of the cellular debris by centrifugation, 10 μl of a 10-mg/ml concentration of pronase was added to the supernatant and incubated for 30 min at 65°C to remove any remaining bacterial proteins. The cosmid DNA was precipitated by adding 20 μl of 5 M ammonium acetate and 900 μl of isopropanol–1 mM phenylmethylsulfonyl fluoride and was then resuspended in 75 μl of TE (pH 8.0). Restriction enzyme digestion of approximately 300 cosmid clones with *Not*I and *Ssp*I demonstrated uniform coverage of pNL1.

Five small insert libraries were constructed to ensure complete representation of pNL1. The construction of a small insert library from a small amount of pNL1 DNA by using a variation on the random PCR amplification method described by Grothues and Tummler (28) has been previously described (101). pNL1 DNA for the second library was obtained from nine cosmid clones whose inserts covered most of pNL1. Approximately 100 μg of cosmid DNA was sheared by sonication (6) to fragments of 0.5 to 1 kb, whose ends were then repaired with exonuclease I and then cloned into pCR-Blunt (Invitrogen, Carlsbad, Calif.). The remaining three libraries were constructed by subcloning pNL1 DNA digested with *Bam*HI, *Eco*RI, or *Pst*I into the cognate sites of pBluescript II SK(+) vector (Stratagene).

Sequencing. Plasmid DNA was purified with the Qiagen QIAwell 96 Ultrawell plasmid kit (Qiagen, Inc., Valencia, Calif.). Cosmid DNA was purified by standard alkaline lysis (55), denatured at room temperature for 10 min in 0.5 M NaOH, neutralized with 1.2 M (final concentration) sodium acetate, and ethanol precipitated. For sequencing reactions, 500 ng of plasmid DNA or 1 μg of cosmid DNA were used for both the cosmid and the plasmid libraries, and primers complementary to the T3 and T7 vector promoter sequences were used to generate the sequence from the insert ends by using dideoxy chain-termination sequencing reactions (83) with Perkin-Elmer/Applied Biosystems Dye Terminators and Ampli-Taq DNA polymerase. For gap closure, primers were designed from the ends of assembled sequences by using Sequencher software (Gene Codes Corp., Ann Arbor, Mich.), MacVector sequence analysis software (Eastman Kodak, Rochester, N.Y.) and Gap4 from the Staden software package. Primers were synthesized by using standard phosphoramidite chemistry on the Applied Biosystems model 392 DNA synthesizer. Sequencing reactions were analyzed with an ABI 377 DNA sequencer.

Sequence assembly. Sequences were assembled by using tools from the Staden Sequence Analysis software package (10). Sequences were initially processed with PREGAP to automate the generation of compressed files for assembly and to mark vector and other non-pNL1 sequences that appear as a result of library construction. These files were assembled and manually edited by using GAP4. Upon completion, a FASTA file was generated for automated sequence analysis.

Automated sequence analysis. Automated sequence analyses were completed by using MAGPIE (multipurpose automated genome project investigation environment) (27). Initial assignments of putative open reading frames (ORFs) were made by using the criteria that (i) the start codon was ATG, GTG, or TTG; (ii) the stop codon was TAA, TAG, or TGA; and (iii) the ORF size was between 150 and 30,000 bp in length. Default MAGPIE settings were used to characterize the 1,346 potential ORFs identified. Data collected was manually surveyed to identify the ORFs most likely to encode a gene. The position of each start codon was estimated manually by identification of the Shine-Dalgarno sites within 15 bp of a potential start codon, the position of upstream genes, and a comparison to the start position of the gene homologues. Protein sequences were aligned with CLUSTALW. ORFs predicted to encode genes were adjusted in size to reflect the manually predicted start position and realigned with GenBank sequences by using BLASTP2 and TBLASTN.

Nucleotide sequence accession number. The sequence of the *S. aromaticivorans* F199 plasmid pNL1 has been deposited with GenBank under accession number AF079317.

Oxygenase activity measurements. Two cosmids were selected from the *E. coli* XL1-Blue large insert library which contained either *bphC* (cosmid 6) or *xylE* (cosmid 18) for oxygenase activity measurements. The sCOS-1 T3 primer target site in these clones is positioned 5,887 and 288 bp upstream of the start codons of *bphC* and *xylE*, respectively. Overnight cultures were harvested and washed in 20 mM sodium phosphate buffer (pH 7.5). After the addition of a final 0.1-mg/ml concentration of DNase, the cells were lysed by the application of two passes through a French pressure cell at 16K lb/in^2. Cellular debris was removed by centrifugation (JA-20; 10,000 rpm for 25 min). The protein content was estimated with a Bio-Rad protein assay kit. Enzymatic production of colored intermediates was measured spectrophotometrically by monitoring the increase in the absorbance at the corresponding wavelength of each *meta* cleavage product formed from the following substrates: catechol, 375 nm; 3-methyl catechol, 388 nm; 4-methyl catechol, 382 nm, and 2,3-dihydroxybiphenyl, 434 nm. The reaction mixture contained 0.1 M sodium phosphate (pH 8) and between 0.5 and 1 mg of cellular lysate. The reaction was initiated by the addition of the appropriate substrate at a final concentration of 0.4 mM for catechol-type substrates or 0.5 mM for 2,3-dihydroxybiphenyl.

Oxidation of 1,2-dihydroxynaphthalene was measured with lysates prepared in 20 mM KH_2PO_4-NaOH (pH 6.2), 1 mM 2-mercaptoethanol, and 10% ethanol. The lysate was treated with oxygen-free ferrous sulfate at a concentration of 0.5 mM for 1 h on ice in an anaerobic glovebox prior to assay. The reaction mixtures contained between 0.5 and 1 mg of cellular lysate in 0.1 M acetic acid-NaOH buffer (pH 5.5) at a total volume of 1.8 ml. The reaction was initiated by the addition of 0.5 μmol of 1,2-dihydroxynaphthalene in 10 μl of tetrahydrofuran. A decrease in oxygen over time as a percentage of the total dissolved oxygen was measured polarographically with a YSI model 5300 dissolved oxygen monitor (Yellow Springs Instrument Co., Yellow Springs, Ohio).

RESULTS

Association of conjugative and biodegradative functions with pNL1. Attempts to transfer native pNL1 to nonaromatic degrading *S. paucimobilis* and *Sphingomonas* sp. strain S88m260 in conjugation experiments were unsuccessful. In order to provide a more robust selectable marker on the plasmid, *S. aromaticivorans* F199 was subjected to transposon mutagenesis. A single mutant, designated F199 tn349, with an altered phenotype that was linked to plasmid function, was identified. *S. aromaticivorans* F199 tn349 was unable to convert indole to indigo, a property commonly affiliated with oxygenases, including styrene monooxygenase (71), xylene monooxygenase (11), toluene dioxygenase (37), and naphthalene dioxygenase (13). This mutant was also unable to grow on naphthalene or biphenyl and could not produce orange metabolites from dibenzothiophene or yellow metabolites from fluorene or biphenyl. It retained its ability to grow on *p*-cresol, *m*-xylene, salicylate, and benzoate and to produce yellow metabolites from catechol.

F199 tn349 was mated with *Sphingomonas* sp. strain S-88 m260 and kanamycin-resistant exconjugants identified. Plasmids were extracted from seven exconjugants and analyzed by gel electrophoresis. Each exconjugant possessed a plasmid the size of pNL1 in addition to the plasmids native to *Sphingomonas* sp. S-88 m260. The degradative properties of the exconjugant were identical to *S. aromaticivorans* F199 tn349, except that they were unable to grow on *p*-cresol. These results suggest that pNL1 encodes the entire *m*-xylene and benzoate catabolic pathways, as well as pathways for growth on salicylate. The inability of *Sphingomonas* S-88 m260 exconjugants to grow on *p*-cresol suggests that either none or only portions of the *p*-cresol degradative pathway were encoded by pNL1.

Nucleotide composition and restriction sites. The circular plasmid pNL1 from *S. aromaticivorans* F199 was 184,457 bp in length. The overall G+C content was 62% and the A+G content was 49%; this value is consistent with an earlier report

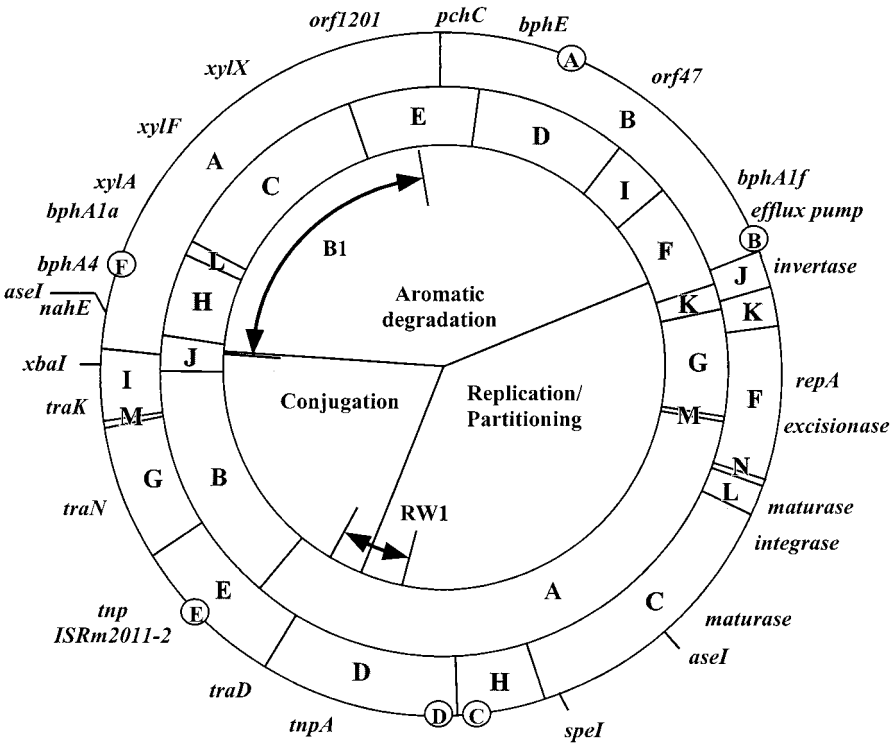


FIG. 1. Physical map of pNL1. Fragments predicted by computational analysis of the pNL1 sequence for *Not*I and *Ssp*I digests are shown in the outer and inner rings, respectively. The positions of selected genes are shown outside the rings, and putative regulators encoded by *orf7* (A), *orf158* (B), *orf569* (C), *orf597* (D), *orf758* (E), and *orf994* (F) are shown in circles on the perimeter of the outer ring. Also depicted are the regions with homology to *S. yanoikuyae* B1 and *Sphingomonas* sp. strain RW1 (bidirectional arrows) and the regions encoding genes associated with aromatic degradation, conjugation, and plasmid partitioning and replication.

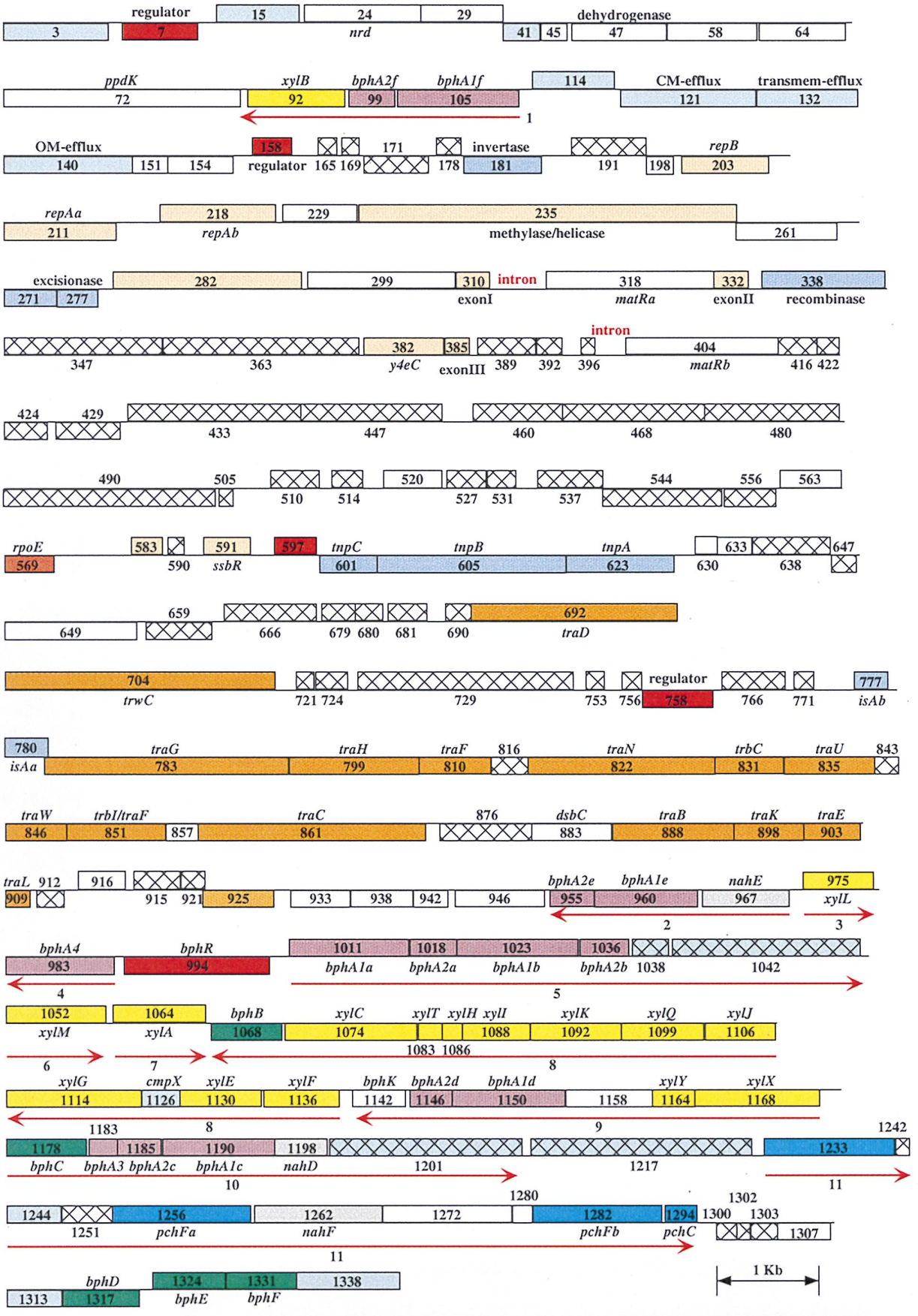
of a G+C content of 62.9 to 65.4 mol% for *S. aromaticivorans* F199 total DNA (5), which suggests that this plasmid was not laterally transferred from another genus.

Examination of the sequence for restriction enzyme sites revealed that there were no *Pme*I sites, single *Xba*I and *Spe*I sites, and two *Ase*I sites in this plasmid; this finding is also consistent with earlier results (86). The reported *Ssp*I digestion pattern included 12 fragments (designated A through L) which were mapped in the order ILFKGABJHCED and sized at 6, 1.8, 12, 2.7, 11, 51, 30, 3.5, 8.5, 25, 14, and 16 kb, respectively. Thirteen *Ssp*I sites were found in the plasmid DNA sequence, with an extra 321 bp *Ssp*I fragment (designated M) present between fragments G and A. The position of fragment L was moved between fragments H and C, which results in an updated *Ssp*I fragment order of IFKGMABJHLCED. The calculated fragment sizes based on DNA sequence analysis are 6,166, 11,510, 2,540, 11,150, 321, 61,361, 26,058, 3,330, 8,596, 1,917, 22,034, 13,655, and 15,717 bp, respectively. The reported *Not*I restriction analysis revealed 14 fragments (designated A through N), and these were mapped in the order BJKLNFCH DEGMIA and sized at 35, 3.5, 3.2, 2.5, 0.6, 13, 24, 7.4, 17, 14, 12, 1.1, 6, and 45 kb, respectively. Fourteen *Not*I restriction sites were also found in the DNA sequence, but the order of LNF was reversed to FNL. The calculated *Not*I fragment sizes for BJKFNLCHDEGMIA are 35,958, 3,266, 2,945, 13,219, 571, 2,649, 24,426, 7,682, 17,289, 13,863, 11,888, 836, 5,734, and 44,259 bp, respectively. Data from the original restriction analyses were reassessed and found to support the changes indicated by the sequence analysis. Agreement between the calculations of restriction fragment sizes and the order based on DNA sequence and agarose gel analyses indicates that the sequence assembly is robust. A revised physical map is depicted in Fig. 1.

ORF analysis. A graphical representation of the ORFs predicted to encode genes is depicted in Fig. 2. Similarities between representative homologs and pNL1 ORFs are detailed in Table 1. Based on gene homolog analysis, functions associated with the plasmid were divided into three general categories: catabolism of aromatic compounds, plasmid replication and partition, and conjugation. The genes associated with each of these general categories are discussed in detail below, with emphasis on those associated with the catabolism of aromatic compounds.

Ring-hydroxylating dioxygenases. Seven homologs to both large and small substrate binding components of ring hydroxylating dioxygenases were identified (*bphA1*[*a–f*]-*bphA2*[*a–f*]). Only a single homolog for each the ferredoxin (*bphA3*) and the ferredoxin reductase (*bphA4*) components were found. A homolog to XylZ, possessing both ferredoxin and ferredoxin reductase domains, was not found. The Cys-X_1-His-X_{17}-Cys-X_2-His Rieske-type (2Fe-2S) cluster binding site motif is conserved in each of the seven pNL1-encoded large oxygenase components. All except BphA1e also possess a potential mononuclear non-heme iron coordination site consensus sequence, E-$X_{3/4}$-D-X_2-H-$X_{4/5}$-H, defined by Jiang et al. (38). In BphA1e (*S. aromaticivorans* F199 and *S. yanoikuyae* B1) and OrfG1 from *Sphingomonas* sp. RW1 (1), two aspartate residues are separated by only one amino acid (D-X_1-D-X_2-H-X_4-H). It is unclear whether these latter proteins can accommodate the Fe(II) ligand.

A dendrogram of representative protein homologs for small- and large-ring hydroxylating oxygenase components is depicted



in Fig. 3. On pNL1, the gene homologs for the large and small oxygenase subunits are all located in pairs. Members of each pair cluster with large and small subunits of similar oxygenases from the same bacterium rather than from dissimilar organisms, suggesting that members of a pair function as components of the same enzyme. For example, XylX and XylY from pNL1 cluster with the XylX and XylY benzoate dioxygenase large and small components, respectively, from *S. yanoikuyae* B1 rather than with components from distinct enzymes or organisms. The distribution of these gene pairs on pNL1 suggests that they are part of at least six different operons. The absence of terminator-like sequences between *bphA2a* and *bphA1b* suggests that the *bphA1a-bphA2b* and *bphA1b-bphA2b* pairs are cotranscribed. These findings suggest that most or all paired binding components are part of the same oxygenase enzyme complex.

Biphenyl, naphthalene, and *m*-xylene catabolic gene homologs. Homologs to genes associated with *m*-xylene, biphenyl, and naphthalene model degradative pathways were identified on pNL1 (Fig. 4). Based on gene spacing and the occurrence of terminator-like structures, we predict that genes encoding enzymes associated with these pathways are distributed among at least 11 transcriptional units (Fig. 2).

Critical analysis of the pNL1 sequences suggests that some genes from both the *m*-xylene and the naphthalene degradative pathways are required for complete degradation of biphenyl. Only a single homolog to dihydrodiol dehydrogenase (*orf1068*; *bphB*) and ring cleavage oxygenase (*orf1178*; *bphC*) are found on pNL1. These enzymes catalyze the second and third steps, respectively, in naphthalene and biphenyl degradation and typically exhibit broad substrate specificity (Fig. 3) (31, 80). Results from analysis of pNL1 subclones encoding *bphC* or *xylE* confirmed that *bphC* is capable of oxidizing both 1,2-dihydroxynaphthalene and 2,3-dihydroxybiphenyl (Table 2). This result further supports the hypothesis that the first three steps in the degradation of both biphenyl and naphthalene are catalyzed by a single set of enzymes whose components are encoded by genes on four different putative transcriptional units: 1 (*bphA1e* and *bphA2e*), 4 (*bphA4*), 8 (*bphB*), and 10 (*bphC* and *bphA3*).

The BphC degradative product is converted by the BphD hydrolase to *cis*-2-hydroxy-penta-2,4-dienoate (2HPDA) and benzoate. The 2HPDA intermediate can then be converted by the products of *bphE* and *bphF* to pyruvate and acetaldehyde. The benzoate intermediate can presumably be further metabolized via the *m*-xylene degradative pathway enzymes. Degradation of benzoate requires genes encoded by putative transcripts 3 (*xylL*) and 9 (*xylXY*), as well as continued expression of genes on putative transcripts 1 (*bphA1e*, *bphA2e*, and *xylB*), 4 (*bphA4*), 8 (*xylFEGJQKIHTC*), and 10 (*bphA3*). This brings the total number of required putative transcripts to eight for the complete metabolism of biphenyl. Two putative transcripts are required for biphenyl degradation alone (*bphD-orf1313-orf130-7orf1303* and *bphE-bphF-orf1338*), three are required for both biphenyl and *m*-xylene degradation (transcripts 3, 8, and 9), and three are required for *m*-xylene, biphenyl, and naphthalene degradation (transcripts 1, 4, and 10).

In addition to benzoate pathway-encoding transcripts 3 and 9, degradation of *m*-xylene requires the expression of genes encoded on transcripts 6 (*xylM*) and 7 (*xylA*), as well as transcript 1 (*xylB*), bringing the total to five transcripts necessary for *m*-xylene degradation. Metabolism of naphthalene to salicylate requires expression of genes from six transcripts, including those also required by biphenyl (1, 4, 8, and 10) and transcripts 2 (*nahE*) and 11 (*nahF*). Further metabolism of salicylate to catechol is typically catalyzed by salicylate hydroxylase (NahG). However, no *nahG* homolog was found on pNL1, suggesting that either salicylate degradation does not proceed through a catechol intermediate or that a novel pNL1-encoded enzyme is responsible for converting salicylate to catechol.

***p*-Cresol catabolic gene homologs.** The prototypical toluene catabolic pathway from P. mendocina KR1 proceeds through a *p*-cresol intermediate and involves the action of three enzymes to generate protocatechuate from *p*-cresol (99). The conversion of *p*-cresol to *p*-hydroxybenzaldehyde is a two-step process that is catalyzed by a two-component *p*-cresol methylhydroxylase. Two gene homologs, designated *pchFa* and *pchFb*, to the *p*-cresol methylhydroxylase flavoprotein component and one cytochrome subunit homolog, *pchFc*, are found on a single cluster in pNL1 (Fig. 2). Subsequent conversion of *p*-hydroxybenzaldehyde to *p*-hydroxybenzoate in *P. mendocina* KR1 is catalyzed by *p*-hydroxybenzaldehyde dehydrogenase. NahF clusters most closely with *p*-hydroxybenzaldehyde dehydrogenase (unpublished results; GenBank no. 995954) and presumably is capable of catalyzing the aldehyde dehydrogenase reactions in both the naphthalene and the *p*-cresol pathways. Homologs to known proteins responsible for catabolism of *p*-hydroxybenzoate were not found on pNL1.

Membrane proteins. Several pNL1 ORFs, distributed among the catabolic genes, have extensive homology to bacterial efflux pump genes. Typical efflux pump transport systems are composed of a cytoplasmic and outer membrane protein and a membrane fusion protein (74). Sequence similarity analysis suggests that ORF140, ORF132, and ORF121 encode the outer membrane, membrane fusion, and cytoplasmic membrane components of a novel efflux pump. Psort analysis of ORF140 identified a possible signal cleavage site and supports its localization in the outer membrane. Tmpred analysis of the putative cytoplasmic component of the pNL1 encoded efflux pump, ORF121, predicts the presence of 14 transmembrane helices, a feature common to Family 1 transport exporters such as EmrB (73). A signature for the *lysR* family of regulatory proteins is present at the N terminus of ORF121 and, like ORF140, it has a membrane lipoprotein lipid A signature.

Two additional putative inner membrane proteins whose function may be related to efflux pumps were also found. The deduced protein sequence of *orf114*, which is immediately downstream of the pNL1-encoded efflux pump cluster, is predicted to have 10 transmembrane domains. It has a lipocalin signature (PS00213) that is characteristic of proteins which transport small hydrophobic molecules such as steroids, bilins, retinoids, and lipids. ORF3 is predicted to have 11 transmembrane domains and is similar to proteins belonging to the drug resistance translocase subfamily BCR/CMLA.

Nine ORFs that cluster among the pNL1-encoded aromatic

FIG. 2. Graphical depiction of ORFs. Putative ORFs are depicted as boxes placed either above (frames 1 to 3) or below (frames 4 to 6) the axis. The numbers correspond to MAGPIE number assignments for each of 1,394 potential ORFs found on pNL1. Red arrows indicate putative transcripts (1 to 11) which encode enzymes involved in biphenyl, *m*-xylene, and naphthalene degradation. Colors are used to highlight putative regulators (red); aromatic oxygenase subunits (purple); enzymes in biphenyl (green), naphthalene (gray), *m*-xylene (yellow), and *p*-cresol (turquoise) pathways; possible aromatic transport proteins (light blue); transposases and recombinases (medium blue); plasmid partitioning and replication (light orange); and conjugative genes (dark orange). Hatched boxes depict ORFs with no detected homologs outside pNL1.

TABLE 1. Homology between pNL1 proteins and representative homologs

pNL1	Functional description of closest relative	pNL1[a] (alignment region)	Homolog[a] (alignment region)	% Identity[b]	Source	GenBank access no.[c]
orf003	Bcr; sulfonamide-bicyclomycin resistance protein	394 (1–378)	396 (14–390)	25	*E. coli* K-12	584831
orf007	CatR; putative regulator of catechol degradative operon	283 (33–283)	256 (3–256)	26	*R. opacus* 1CP	2398776
orf015	DesA; fatty acid desaturase	309 (27–283)	287 (22–268)	29	*C. caldarium* RK1 chloroplast	2465769
nrd	Nrd; aerobic-microaerobic growth with nitrates	439 (3–438)	444 (4–442)	69	*B. japonicum* USDA 110	2961296
orf029	YkpB; conserved hypothetical protein	310 (3–303)	303 (2–297)	20	*B. subtilis* 168	2633815
	ORF58; conserved hypothetical protein	310 (4–294)	337 (7–322)	28	*S. aromaticivorans* F199 (pNL1)	3378270
orf041	ORF1; conserved hypothetical protein	136 (1–131)	139 (1–131)	51	*B. cepacia* LB400	586492
orf045	YtvB; hypothetical protein	94 (18–52)	111 (65–99)	34	*B. subtilis*	2293180
orf047	ImdH; probable tartrate dehydrogenase	355 (1–353)	361 (2–360)	60	*E. coli* K-12	3024767
	DlpA; isocitrate-isopropylmalate dehydrogenase family	355 (8–350)	615 (9–362)	32	*L. pneumophila*	2497266
orf058	YkpB; thiamin biosynthesis-like hypothetical protein	337 (11–335)	301 (7–295)	28	*P. horikoshii* OT3	3257813
orf064	DapA; putative dihydropicolinate synthase	320 (18–317)	289 (4–285)	25	*A. fulgidus*	2649690
ppdK	PpdK; pyruvate phosphate dikinase	888 (21–885)	874 (16–871)	57	*C. symbiosum*	1084302
xylB	XylB; aryl alcohol dehydrogenase	365 (1–365)	366 (1–365)	51	*P. putida* (pWW53)	731175
bphA2f	ORFG6; small subunit	174 (23–173)	176 (22–175)	38	*Sphingomonas* sp. strain RW1	3618275
bphA1f	PahA3; large subunit component naphthalene dioxygenase	459 (7–436)	449 (8–439)	44	*P. aeruginosa* PaK1	1255669
orf114	AF010496; conserved hypothetical protein	309 (17–306)	289 (9–295)	32	*R. capsulatus* SB1003	3128363
orf121	VceB; putative multidrug resistance pump gene 2	509 (21–502)	511 (22–504)	43	*V. cholerae* 569B	2815578
orf132	VceA; putative multidrug resistance pump gene 1	381 (19–380)	395 (16–376)	46	*V. cholerae* 569B	2815577
orf140	OprJ; outer membrane component of multidrug resistance efflux pump	483 (12–473)	479 (12–468)	30	*P. aeruginosa* PAO	2497711
orf151	ORF151; conserved hypothetical protein	137 (25–135)	151 (25–145)	32	*H. halobium*	141254
orf154	BpoA; conserved hypothetical protein	247 (2–242)	261 (21–259)	43	*M. tuberculosis* H37Rv	2104396
orf158	YybA; hypothetical regulator, Marr family	148 (24–148)	150 (11–135)	25	*B. subtilis* 168	2636618
orf181	Y4cG; probable DNA invertase	293 (1–292)	305 (16–303)	45	*Rhizobium* sp. strain NGR234(pNGR234α)	2497425
orf198	Y4dW; conserved hypothetical protein	103 (21–85)	204 (91–155)	36	*Rhizobium* sp. strain NGR234(pNGR234α)	2496610
repB	QsopB; *trans*-acting replication protein	327 (54–299)	334 (62–296)	28	*C. burnetii* (QpH1)	1075479
repAa	RepA; DNA replicase	420 (22–412)	405 (2–402)	32	*A. tumefaciens* (pTiB6S3)	95114
repAb	RepA; DNA replicase	434 (52–312)	307 (29–284)	28	*R. marinus* R-21(pRM21)	527643
orf229	ORF248; conserved hypothetical protein	280 (103–174)	248 (165–239)	41	*C. burnetii* (QpDV)	1075477
orf235	R31180_1; conserved hypothetical protein	1425 (375–1146)	1384 (197–1015)	31	*H. sapiens*	3399676
	HsdM; StySKI methylase	1425 (121–298)	493 (167–373)	23	*S. enterica*	1841495
orf261	AF056496; conserved hypothetical protein	379 (100–288)	252 (93–252)	22	*H. pylori* OU145B(pHP0100)	3046602
orf271	Rv3749c; hypothetical protein	196 (99–177)	169 (60–138)	35	*M. tuberculosis* H37Rv	2960173
orf277	Rv3750c; hypothetical protein similar to excisionases	152 (53–144)	130 (23–114)	38	*M. tuberculosis* H37Rv	2960174
orf282	ORF1; 20.4-kDa plasmid stabilizing protein	705 (119–288)	180 (1–168)	29	*R. meliloti* GR4	46317
	ORF2; 38.5-kDa plasmid stabilizing protein	705 (310–624)	356 (21–324)	26	*R. meliloti* GR4	46318
	AF053947; conserved hypothetical protein	705 (4–632)	669 (21–624)	27	*Y. pestis* (pMT1)	2996364
orf299	AF0450; hypothetical signal transducing His kinase	555 (30–128)	329 (229–329)	31	*A. fulgidus*	2650174
orf310	ORF382; conserved hypothetical protein	127 (5–109)	300 (9–118)	44	*S. aromaticivorans* F199(pNL1)	3378306
	Y4eC; conserved hypothetical protein	127 (8–108)	307 (9–118)	35	*Rhizobium* sp. strain NGR234(pNGR234α)	2496614
matRa	MatRb; putative maturase-related protein	633 (45–611)	571 (3–570)	52	*S. aromaticivorans* F199(pNL1)	3378311
	MatR; maturase-related protein	633 (26–601)	599 (26–594)	38	*L. lactis* MG1363	1296829
orf332	Y4eC; conserved hypothetical protein	130 (13–115)	307 (137–235)	37	*Rhizobium* sp. strain NGR234(pNGR234α)	2496614
	ORF382; conserved hypothetical protein	130 (13–118)	300 (134–235)	37	*S. aromaticivorans* F199(pNL1)	3378306

Continued on following page

TABLE 1—*Continued*

pNL1	Functional description of closest relative	pNL1[a] (alignment region)	Homolog[a] (alignment region)	% Identity[b]	Source	GenBank access no.[c]
orf338	XerD; putative phage-type integrase-recombinase	463 (202–459)	316 (143–346)	23	*M. leprae*	2065218
orf382	Y4eC; conserved hypothetical protein	300 (6–292)	307 (7–295)	56	*Rhizobium* sp. strain NGR234(pNGR234α)	2496614
orf385	Y4eC; conserved hypothetical protein	90 (14–90)	307 (225–307)	36	*Rhizobium* sp. strain NGR234(pNGR234α)	2496614
	ORF382; conserved hypothetical protein	90 (22–80)	300 (235–293)	32	*S. aromaticivorans* F199(pNL1)	3378306
matRb	MatRa; putative maturase-related protein	571 (3–570)	633 (45–611)	52	*S. aromaticivorans* F199(pNL1)	338301
	MatR; maturase related protein	571 (6–561)	599 (48–594)	36	*L. lactis* MG1363	1296829
orf447	ORF480; hypothetical protein	533 (18–386)	507 (14–368)	29	*S. aromaticivorans* F199(pNL1)	3378320
orf480	ORF447; hypothetical protein	507 (14–368)	533 (18–386)	29	*S. aromaticivorans* F199(pNL1)	3378317
orf520	Rv3226c; conserved hypothetical protein	220 (20–189)	252 (1–200)	28	*M. tuberculosis* H37Rv	2072693
orf563	HglK; protein involved in glycolipid localization	229 (12–206)	245 (52–240)	33	*Synechocystis* sp. strain PCC6803	1653115
orf569	RpoE; RNA polymerase sigma-E factor (sigma 24)	181 (4–170)	202 (11–192)	26	*E. coli* K-12	987648
orf583	ORFL1; conjugal transfer region	114 (16–113)	152 (69–151)	43	*A. tumefaciens* (Ti)	ATU43674
ssbR	Ssb; homolog of single-strand binding protein	111 (3–111)	113 (4–104)	30	*E. aerogenes* (R751)	1572546
orf597	Ros; repressor of plant conjugal transfer genes	159 (2–128)	142 (9–137)	53	*A. tumefaciens* (Ti)	1710641
tnpC	TnpC; type Tn*554* transposase homolog	218 (18–109)	125 (17–106)	23	*S. aureus*	136146
tnpB	TnpB; Tn*554* transposase	710 (373–622)	630 (336–578)	30	*S. aureus*	135956
tnpA	TnpA; type Tn*554* transposase homolog	404 (24–374)	361 (16–347)	31	*S. aureus*	135955
orf630	PHS027; conserved hypothetical protein	81 (7–68)	74 (6–66)	35	*P. horikoshii* OT3	3257321
orf633	Rv2759c; conserved hypothetical protein	129 (4–108)	131 (1–115)	27	*M. tuberculosis* H37Rv	2624281
orf649	ORF2; putative 27.6-kDa protein	495 (261–440)	236 (31–212)	27	*H. pylori* (pHPM180)	530827
	Tca; DNA polymerase	495 (31–262)	834 (276–528)	28	*T. aquaticus caldophilus* GK24	2506365
traD	TraD; membrane protein involved in DNA export	776 (237–726)	738 (183–677)	30	*E. coli* (R100)	136176
trwC	TrwC; DNA helicase	1013 (2–955)	966 (4–929)	31	*E. coli* (R388)	1084124
orf758	ORF4; hypothetical regulatory protein	266 (11–201)	251 (7–209)	24	*Acinetobacter* sp. strain ADP1	2271498
isAb	ORFA; transposase part A ISRm2011-2	128 (1–126)	135 (1–126)	40	*R. meliloti* 2011	1583631
isAa	ORFB; transposase part B ISRm2011-2	157 (1–155)	201 (45–199)	50	*R. meliloti* 2011	1583632
traG	TraG; pilus biosynthesis-mating aggregate stabilization	912 (1–830)	938 (1–802)	21	*E. coli* K-12 (F)	464922
traH	TraH; F pilus assembly	484 (21–469)	458 (10–443)	32	*E. coli* K-12 (F)	464925
traF	TraF; F pilus assembly	270 (48–263)	247 (23–241)	27	*E. coli* (R100-1)	2226078
traN	TraN; stabilization of mating pairs	704 (396–695)	602 (293–593)	29	*E. coli* K-12 (F)	730985
trbC	TrbC; F pilus assembly	254 (99–221)	212 (74–206)	27	*E. coli* (R100-1)	2226075
traU	TraU; conjugal DNA transfer-F pilus assembly	340 (19–340)	330 (13–330)	44	*E. coli* K-12 (F)	136155
traW	TraW; F pilus assembly	231 (12–210)	210 (5–204)	36	*E. coli* (R100-1)	2226072
orf851	TraF; F pilus assembly	368 (209–340)	177 (9–141)	32	*E. coli* (RK2)	549119
	TrbI; influences F pilus outgrowth and retraction	368 (85–150)	128 (56–125)	31	*Escherichia coli* (R100-1)	2226071
traC	TraC; F pilus assembly	856 (5–833)	875 (24–851)	25	*E. coli* K-12 (F)	136145
dsbC	DsbC; protein disulfide-isomerase	298 (182–290)	235 (110–223)	32	*E. coli*	96548
traB	TraB; F pilus assembly	458 (36–453)	475 (4–406)	26	*E. coli* (ColB2)	730990
traK	TraK; F pilus assembly	264 (32–263)	242 (8–241)	20	*E. coli* (ColB2)	1293089
traE	TraE; F pilus assembly	207 (19–183)	187 (12–186)	22	*S. typhi* (pED208)	136180
traL	TraL; F pilus assembly	94 (1–79)	91 (8–86)	31	*E. coli* (ColB2)	136149
orf916	Phbn040; hypothetical regulator	177 (80–156)	158 (26–93)	29	*P. horikoshii* OT3	3130383
orf925	YjbJ; conserved hypothetical protein similar to lytic transglycolases	265 (103–217)	181 (63–177)	47	*B. subtilis*	2633511
orf933	MTH888; hypothetical protein	222 (7–203)	220 (22–214)	32	*M. thermoautotrophicum*	2621982
	ORF938; conserved hypothetical protein	222 (2–216)	232 (9–226)	34	*S. aromaticivorans* F199(pNL1)	3378388
	DlpA; isocitrate-isopropylmalate dehydrogenase family	222 (6–219)	615 (395–611)	36	*L. pneumophila*	2497266
orf938	MTH888; hypothetical protein	232 (25–217)	220 (33–217)	34	*M. thermoautotrophicum*	2621982
	DlpA; isocitrate-isopropylmalate dehydrogenase family	232 (1–214)	615 (374–595)	29	*L. pneumophila*	2497266

Continued on following page

TABLE 1—*Continued*

pNL1	Functional description of closest relative	pNL1[a] (alignment region)	Homolog[a] (alignment region)	% Identity[b]	Source	GenBank access no.[c]
orf942	PH0854; hypothetical protein	130 (16–125)	137 (27–135)	32	*P. horikoshii* OT3	3257265
orf946	Slr0619; hypothetical protein	336 (53–331)	348 (66–334)	23	*Synechocystis* sp. strain PCC6803	1208451
bphA2e	BphA2e; small subunit component of a dioxygenase	165 (1–165)	165 (1–165)	72	*S. yanoikuyae* B1	
bphA1e	BphA1e; large subunit component of a dioxygenase	391 (1–391)	391 (1–391)	90	*S. yanoikuyae* B1	
nahE	NahE; 2-hydroxy-benzalpyruvate aldolase	328 (1–327)	333 (1–327)	83	*S. yanoikuyae* B1	
xylL	XylL; *cis*-(methyl)benzoate dihydrodiol dehydrogenase	262 (7–262)	263 (8–263)	76	*S. yanoikuyae* B1	
bphA4	BphA4; reductase component of a dioxygenase	408 (1–408)	407 (1–407)	72	*S. yanoikuyae* B1	
bphR	BphR; regulatory protein	546 (1–542)	543 (1–541)	61	*S. yanoikuyae* B1	
bphR	BphR; regulatory protein	546 (1–542)	543 (1–541)	61	*S. yanoikuyae* B1	
bphA1a	BphA1a; large subunit component of a dioxygenase	450 (5–450)	448 (3–448)	82	*S. yanoikuyae* B1	
bpha2a	BphA2a; small subunit component of a dioxygenase	175 (1–175)	175 (1–175)	65	*S. yanoikuyae* B1	
bphA1b	BphA1b; large subunit component of a dioxygenase	455 (1–455)	455 (1–455)	89	*S. yanoikuyae* B1	
bphA2b	BphA2b; small subunit component of a dioxygenase	183 (1–183)	183 (1–183)	80	*S. yanoikuyae* B1	
orf1038	ORF1217; conserved hypothetical protein	135 (1–135)	832 (1–135)	79	*S. aromaticivorans* F199(pNL1)	3378429
orf1042	ORF1217; conserved hypothetical protein	705 (1–705)	832 (152–832)	72	*S. aromaticivorans* F199(pNL1)	3378429
xylM	XylM; reductase component of xylene monooxygenase	367 (1–363)	375 (1–371)	67	*S. yanoikuyae* B1	
xylA	XylA; oxygenase component of xylene monooxygenase	346 (1–343)	346 (1–343)	63	*S. yanoikuyae* B1	
bphB	BphB; *cis*-biphenyl dihydrodiol dehydrogenase	266 (1–265)	266 (1–265)	80	*S. yanoikuyae* B1	
xylC	XylC; benzylaldehyde dehydrogenase	501 (1–501)	499 (1–499)	89	*S. yanoikuyae* B1	
xylT	XylT; ferredoxin, plant-type	93 (3–92)	92 (2–91)	68	*S. yanoikuyae* B1	
xylH	XylH; 4-oxalocrotonoate tautomerase	80 (1–69)	70 (1–69)	79	*S. yanoikuyae* B1	
xylI	XylI; 4-oxalocrotonate decarboxylase	256 (1–256)	256 (1–256)	85	*S. yanoikuyae* B1	
xylK	XylK; 4-hydroxy-2-oxovalerate aldolase	343 (1–342)	343 (1–342)	90	*S. yanoikuyae* B1	
xylQ	XylQ; acetaldehyde dehydrogenase (acylating)	312 (1–312)	312 (1–312)	82	*S. yanoikuyae* B1	
xylJ	XylJ; 2-hydroxypent-2,4-dienoate hydratase	264 (1–264)	264 (1–264)	91	*Pseudomonas* sp. strain DJ77	2315998
xylG	XylG; 2-hydroxymuconic semialdehyde dehydrogenase	505 (12–505)	494 (1–494)	96	*Sphingomonas* sp. strain HV3(pSKY4)	2293079
cmpX	CmpX; conserved hypothetical protein	143 (1–143)	143 (1–143)	97	*Sphingomonas* sp. strain HV3(pSKY4)	2293078
xylE	XylE; catechol 2,3-dioxygenase	307 (1–307)	307 (1–307)	95	*Sphingomonas* sp. strain HV3(pSKY4)	2293077
xylF	XylF; 2-hydroxymuconic semialdehyde hydrolase	283 (1–282)	292 (1–282)	93	*Sphingomonas* sp. strain HV3(pSKY4)	2293079
bphK	BphK; glutathione *S*-transferase	201 (1–201)	201 (1–201)	97	*S. yanoikuyae* B1	
bphA2d	BphA2d; small subunit component of a dioxygenase	162 (7–162)	159 (4–159)	76	*S. yanoikuyae* B1	
bphA1d	BphA1d; large subunit component of a dioxygenase	426 (1–426)	424 (1–424)	80	*S. yanoikuyae* B1	
Orf1158	ORF2; conserved hypothetical protein	330 (9–327)	330 (7–325)	87	*S. yanoikuyae* B1	
xylY	XylY; small subunit component of benzene dioxygenase	164 (1–164)	164 (1–164)	69	*S. yanoikuyae* B1	
xylX	XylX; large subunit component of benzene dioxygenase	471 (1–462)	470 (1–462)	73	*S. yanoikuyae* B1	
bphC	BphC; 2,3-dihydroxybiphenyl 1,2-dioxygenase	299 (1–298)	299 (1–298)	87	*S. yanoikuyae* B1	115106
bphA3	PhnR; ferredoxin component of phenanthrene dioxygenase	108 (1–108)	108 (1–108)	97	*Pseudomonas* sp. strain DJ77	2316014
bphA2c	BphA2c; small subunit component of a dioxygenase	162 (1–162)	162 (1–162)	67	*S. yanoikuyae* B1	
bphA1c	BphA1c; large subunit component of a dioxygenase	420 (18–415)	419 (19–415)	79	*S. yanoikuyae* B1	
nahD	NahD; 2-hydroxychromene-2-carboxylate isomerase	197 (1–197)	197 (1–197)	79	*S. yanoikuyae* B1	
orf1201	ORF1217; conserved hypothetical protein	726 (10–726)	832 (19–832)	26	*S. aromaticivorans* F199(pNL1)	3378429

Continued on following page

TABLE 1—*Continued*

pNL1	Functional description of closest relative	pNL1[a] (alignment region)	Homolog[a] (alignment region)	% Identity[b]	Source	GenBank access no.[c]
orf1201	ORF1217; conserved hypothetical protein	726 (10–726)	832 (19–832)	26	*S. aromaticivorans* F199(pNL1)	3378429
orf1217	ORF1042; conserved hypothetical protein	832 (152–832)	705 (1–705)	72	*S. aromaticivorans* F199(pNL1)	3378402
gabD	GabD; succinic semialdehyde dehydrogenase	490 (8–484)	482 (3–482)	57	*E. coli* K-12	120777
orf1244	YclB; phenylacrylic acid decarboxylase homolog	204 (3–190)	204 (14–200)	50	*B. subtilis* 168 trpC2	1805434
pchFa	PchF; flavoprotein subunit, *p*-cresol methyl-hydroxylase	519 (5–479)	530 (9–487)	42	*P. putida* NCIMB 9866	294369
nahF	Ydh; vanillin: NAD^+ oxidoreductase	479 (1–473)	482 (1–475)	41	*P. fluorescens* AN103	3116018
orf1272	YclC; conserved hypothetical protein	486 (17–457)	473 (7–448)	47	*B. subtilis* 168 trpC2	1805435
orf1280	YclD; conserved hypothetical protein	74 (6–62)	229 (4–57)	32	*B. subtilis* 168 trpC2	1805436
pchFb	PchFa; flavoprotein subunit, *p*-cresol methyl-hydroxylase	488 (2–486)	519 (9–504)	37	*S. aromaticivorans* F199(pNL1)	3378433
	PchF; flavoprotein subunit, *p*-cresol methyl-hydroxylase	519 (39–463)	530 (57–487)	32	*P. putida* NCIMB 9866	294369
pchC	PchC; cytochrome *c* subunit, *p*-cresol methyl-hydroxylase	121 (10–115)	101 (10–101)	29	*P. putida* NCIMB 9866	294349
orf1300	FadE19; acyl-CoA dehydrogenase	76 (12–83)	394 (53–124)	36	*M. tuberculosis* H37Rv	2113933
orf1303	Rv3548c; hypothetical dehydrogenase	123 (1–49)	304 (210–256)	32	*M. tuberculosis* H37Rv	1666112
orf1307	Rv0148; steroid dehydrogenase	200 (9–170)	286 (9–169)	66	*M. tuberculosis* H37Rv	1877273
	Rv3548c; hypothetical dehydrogenase	200 (5–169)	304 (4–178)	42	*M. tuberculosis* H37Rv	1666112
orf1313	NK[a]; hypothetical protein	201 (42–182)	159 (5–142)	30	*P. aeruginosa* PAO1	77647
bphD	BphD[d]; 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase	286 (1–283)	285 (1–284)	60	*Rhodococcus* sp. strain RHA1 (390-kb plasmid)	1906778
bphE	BphE[d]; 2-hydroxypenta-2,4-dienoate hydratase	271 (11–270)	267 (7–266)	66	*Rhodococcus* sp. strain RHA1 (390-kb plasmid)	1906779
bphF	BphF[d]; 4-hydroxy-2-oxovalerate aldolase	262 (1–243)	258 (1–243)	55	*Rhodococcus* sp. strain RHA1 (390-kb plasmid)	1906780
orf1338	BaiF; conserved hypothetical protein	388 (18–281)	381 (5–273)	38	*E. coli*	1799782

[a] NK, no gene name designated.
[a] Values in parentheses describe the portion of protein used in the alignment.
[b] Percentage of amino acids in aligned region that are identical when the sequences are aligned with BLASTP2.
[c] Accession number 1163200 describing ORFL has been removed from GenBank. The accession number listed refers to the DNA sequence from which the corresponding ORF can be derived. Sequences for *S. yanoikuyae* B1 genes were derived from the DNA sequences provided by Gerben Zylstra where not available in GenBank.
[d] The gene designation refers to those assigned by Masai et al. (56). In GenBank the same genes are designated *etbDEF*.

catabolic genes are not homologous to functionally characterized genes and are predicted to reside in the membrane. ORF1038 and ORF1042 have extensive homology to the N terminus and the C terminus, respectively, of ORF1217 and to a lesser extent to those of ORF1201. Read-through of the ORF1038 stop codon would result in a single ORF encompassing both ORF1038 and ORF1042. Reanalysis of the raw sequence confirmed that the stop codon separating ORF1038 and ORF1042 was not due to sequence error. The results of Psort analyses predict that ORF1038 (or a fusion of ORF1038 and ORF1042) and ORF1201 are outer membrane proteins and that ORF1217 is localized to the inner (cytoplasmic) membrane.

Homologs to three of the remaining proteins cluster with genes associated with aromatic catabolism and are predicted to reside in the inner membrane. Like the pNL1-encoded *cmpX*, its three homologs from *Pseudomonas* sp. strain DJ77 (unpublished results; GenBank no. 2316006), *Sphingomonas agrestis* HV3 (105), and *S. yanoikuyae* B1 (45) are all found downstream of a catechol 2,3-dioxygenase. Homologs to ORF41 are clustered with components of ring-hydroxylating dioxygenases in *C. oligotrophicus* RB1 (95), *Pseudomonas* sp. strain KKS102 (43), *Ralstonia eutropha* A5 (58), *P. fluorescens* (unpublished results; GenBank no. 1256705), *Pseudomonas* sp. strain JR1 (75), *Comamonas testosteroni* B-356 (90), *P. pseudoalcaligenes* KF707 (92), and *B. cepacia* LB400 (14). ORF1338 is homologous to a partial ORF that is expressed bidirectionally from the *Acinetobacter calcoaceticus cis*,*cis*-muconate catechol *ortho*-cleavage pathway transport protein (100) and to a partial ORF found in a *Bordetella pertussis* gene cluster along with a salicylate hydroxylase homolog (102).

The remaining two membrane proteins, ORF15 and ORF1313, are also predicted to reside in the inner membrane, but they do not align with proteins encoded in other aromatic catabolic gene clusters. ORF15 is predicted to possess five transmembrane domains and is similar to fatty acid desaturases and beta-carotene ketolases. Aligned proteins share three histidine-rich domains with the signatures H-D-X_2-H, H-X_2-H-H and H-X_2-H-H-[LR]-[CWH]-[PV]-X_2-P. ORF1313 has homology to a small hypothetical protein that clusters with regulatory genes that control synthesis of the exopolysaccharide alginate in *P. aeruginosa* (41).

Other ORFs found in the aromatic catabolic region of pNL1. Homologs to two additional ORFs, *bphK* and *orf1158*, also cluster with aromatic catabolic genes, but their function in aromatic catabolism has not been established. BphK is a glu-

tathione *S*-transferase, and its homologs are found in polycyclic aromatic catabolic gene clusters of *Pseudomonas* sp. strain DJ77 (unpublished data), *S. paucimobilis* epa505 (54), *Pseudomonas pseudoalcaligenes* KF707 (49), *Pseudomonas* sp. strain LB400 (34), *S. yanoikuyae* B1 (45), and *Cycloclasticus oligotrophus* RB1 (95). Homologs to ORF1158 are found in the *B. cepacia* Pc701 plasmid-encoded 4-methylphthalate catabolic cluster (82) and in *S. yanoikuyae* B1 (ORF2) (107). The latter ORFs are also similar to the pyridoxal phosphate biosynthetic protein (PdxA).

The eight genes between *orf15* and *xylB* do not align to genes associated with aromatic catabolic gene clusters. ORF24 is similar to the Nrd protein of *Bradyrhizobium japonicum*, whose function is involved in aerobic and microaerobic growth with nitrate as the only nitrogen source. A functional role of ORF24 in microaerobic growth on aromatics is correlates well with earlier findings which demonstrated that *S. aromaticivorans* F199 grew better on aromatic compounds under microaerobic conditions (22). ORF24 is also similar to one member of two component monooxygenases, including nitrilotriacetate monooxygenase (GenBank no. 1119211), dibenzothiophene monooxygenase (GenBank no. 595291), and pristinamycin oxygenase (GenBank no. 940348). A homolog to the second monooxygenase component is not evident on pNL1. Only two of the remaining ORFs in this region, *orf72* (pyruvate, orthophosphate dikinase) and *orf47* (isopropylmalate-tartrate dehydrogenase), show significant homology to functionally characterized genes.

Six new ORFs are linked with aromatic catabolic genes in pNL1 cluster 11, which includes the *p*-cresol catabolic genes and *nahF*. Among these is GabD which, when added to NahF, XylC, XylQ, and XylG, brings the total of pNL1-encoded semialdehyde dehydrogenases to five. ORFS 1244, 1272, and 1280 are similar to first three genes in the *B. subtilis* 168 *yclBCDE* gene cluster (52). ORF1280 contains only enough sequence to align with the first third of YclD and therefore probably does not represent a functional homolog. The remaining two ORFs in the pNL1 cluster 11, ORF1242 and ORF1251, are novel genes.

ORF1307 is encoded in the cluster containing *bphD* and is similar to the N-terminal region of beta-oxoacyl(acyl-carrier protein) reductases, which catalyze the first reduction step in fatty acid biosynthesis, of multifunctional enzymes from eukaryotic organisms that are involved in the beta-oxidation fatty acids and of hypothetical dehydrogenase proteins in *M. tuberculosis*. Within the region of overlap with these three classes of proteins, it has an imperfect (two mismatches) short-chain dehydrogenase-reductase family signature (PS00061). The region of homology with these proteins is extended into the C-terminal region by including the sequence from ORF1303, which can be achieved by a single frame shift. Two additional small ORFs, ORF1302 and ORF1300, are predicted in the remaining gap between *orf1307* and *pchC*. ORF1300 has homology to the N terminus of acyl coenzyme A (acyl-CoA) dehydrogenases, which are also involved in fatty acid oxidation pathways. A frameshift extends the region of homology and adds the acyl-CoA dehydrogenase signature 1 (PS00072). Although additional frameshifts extend the homology to acyl-CoA dehydrogenases even further, the entirety of this region only extends approximately halfway through the homolog sequences.

Plasmid partition-replication. Two RepA-like proteins are encoded by the bidirectionally transcribed genes *orf211* (*repAa*) and *orf218* (*repAb*). Their deduced protein sequences are dissimilar and consequently cluster with different families of RepA proteins. The intergenic space between *repAa* and *repAb* has six 17-bp direct repeats with the consensus sequence 5′-rsCGATGAwyTCvGwkC. Five repeats are located on the strand that encodes *repAb*, and one is on the *repAa*-encoding strand. An additional repeat with this consensus is found at the end of *repAb*. In the intergenic space between *repAa* and *repB* are three inverted repeats with the consensus sequence 5′-AG TTGCCACGTGGCAAC.

Downstream of the *repAb* gene are several additional genes which may be associated with pNL1 replication and partitioning. ORF229 is similar to the C-terminal region of hypothetical proteins associated with *Coxiella burnetii* plasmids and is likely cotranscribed with ORF235, the largest protein encoded on pNL1 (1,425 amino acids). The N-terminal 300-amino-acid portion of ORF235 is similar to methylases and includes an N-6 adenine-specific DNA methylase signature sequence (PS00092) typical of enzymes that methylate the amino group of the C-6 position of adenines. The central region of the protein is similar to several eukaryotic proteins, but the presence of a DEAH-box-family ATP-dependent helicase signature distinguishes it from its eukaryotic homologs. It is likely, therefore, that this novel prokaryotic protein is involved in the DNA methylation and unwinding of pNL1 and possibly other functions related to those catalyzed by its eukaryotic homologs.

ORF282 is homologous to what would be a fusion of the *Rhizobium meliloti* pRmeGr4α plasmid-encoded *orf1* and *orf2* genes, whose products function in plasmid stabilization (57). ORF583 is homologous to members of the thermonuclease family, including the *E. coli* ParB protein that is involved in plasmid partitioning. Like other members of this family, ORF583 has three nuclease active-site residues and is likely localized outside of the cytoplasmic membrane.

Two group II-associated maturases, *matRa* and *matRb*, were found in the pNL1 replication region. No detectable sequence homology was found in the vicinity of *matRb*, but a number of ORFs with homology to portions of TraC proteins encoded by plasmid RP4 and R751 (59) flank *matRa*. The TraC replication primase is required for autonomous replication of the F plasmid in *E. coli*, catalyzing the synthesis of short oligoribonucleotide primers on single-stranded DNA templates. Alternate start sites in these proteins result in expression of N-terminal truncated proteins. The TraC region lost as a result of translation from the internal start sites is homologous to the *Rhizobium* sp. pNGR234*a* plasmid Y4eC protein (23) and pNL1 ORF382. A second ORF382-like protein can be constructed by splicing together the sequences of ORF383 (exon I), ORF332 (exon II), and ORF385 (exon III). Splicing activities associated with *matRa* could presumably splice together exons I and II. However, a means for the addition of exon III, which is separated from exon II by four genes, is less apparent. Except for *matrRb*, genes found between *orf385* and *orf563* do not possess extensive homology to functionally characterized genes.

Integration-recombination-associated genes. pNL1 ORFs predicted to function in DNA integration or recombination include a resolvase family site-specific recombinase (*orf277*), an excisionase (*orf338*), a phage-type integrase-recombinase (*orf338*), and two transposons (*tnpABC* and *isAab*). Since *orf277* and *orf271* are homologous to adjacent genes in *M. tuberculosis* (76), it is probable that both ORFs are functionally related. Sequences from the region encoding *isAab* resemble the ISRm2011-2 transposase of *R. meliloti* (84). Like ISRm2011-2, there is a potential frameshifting window (5′-A AAAAAAG) near the end of IsAa that would allow its fusion with IsAb to form the mature transposase IsA. IsA is flanked by inverted repeat sequences: 5′-CGCGCTAGAGCGGTTTT CGA ending 79 bp upstream of *isAa* and 5′-TCGAAAATCG CTCTAGCGCG starting immediately after the TGA stop

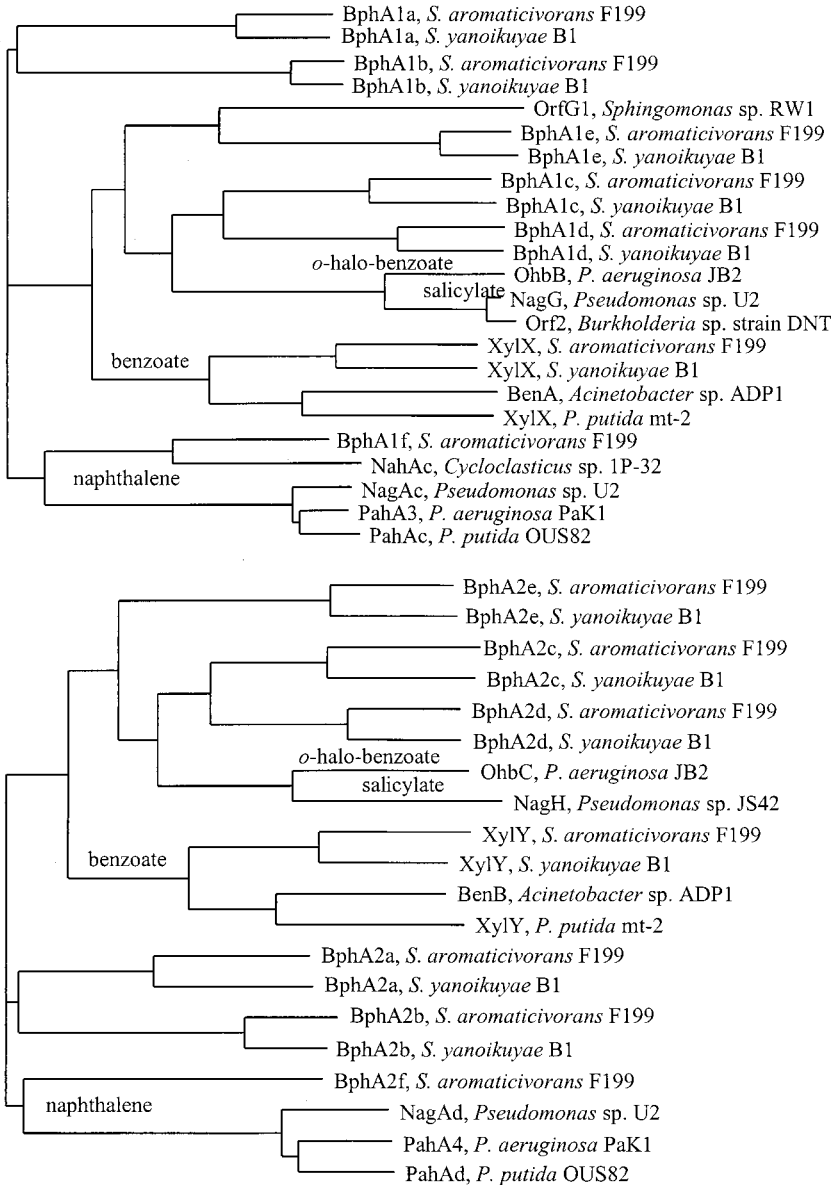


FIG. 3. CLUSTALW dendrogram of representative large (above) and small (below) oxygenase component amino acid sequences. The GenBank accession numbers associated with these sequences are as follows: *Acinetobacter calcoaceticus* ADP1 (2996621-2996622), *P. putida* mt-2 plasmid pWWO (139861-139862), *Pseudomonas* sp. strain U2 plasmid pWWU2 (2828018-2828019, 2828015-2828016), *Sphingomonas* sp. strain RW1 (3618270), *Burkholderia* sp. strain DNT (1477921), *Cycloclasticus* sp. strain 1P032 (3170521), *P. aeruginosa* PaK1 (12555669-1255670), *P. putida* OUS82 (1073044-1073045), and *P. aeruginosa* JB2 (3643998-3643999). Sequences from *S. yanoikuyae* B1 were derived from the dissertation thesis of E. Kim (45).

codon in *isAb*. These repeats are flanked by an upstream CG and a downstream GC putative target of the duplication sequence.

Conjugative genes. Three gene clusters encode homologs to *E. coli* F plasmid genes required for conjugative sex pilus formation and mating-pair stabilization (24). The first cluster includes homologs to the F plasmid genes *traL*, *traE*, *traK*, and *traB*; to *dsbC* (61); and to *S. typhimurium* plasmid R64 *trbB* (26). DsbC catalyzes disulfide bond formation in some periplasmic proteins. Like DsbC, ORF883 possesses an active site typical of disulfide isomerases (FsdfrCgyC) and is predicted to reside in the periplasm. The second cluster includes homologs to *E. coli* F plasmid genes *traC*, *trbI*, *traW*, *traU*, *traN*, *traH*, and *traG*. Homology to the entire TrbI is localized within the N-terminal portion of ORF851. The C-terminal portion is homologous to TraF from *E. coli* RK2 (97) and to various type-I leader peptidases. The third cluster encodes an F plasmid-like TraD and a protein that is similar to TrwC from *E. coli* R388 (53) and to the N-terminal portion of F plasmid TraN. A total of six additional ORFs are also present in these clusters but have no identifiable homolog.

The nucleotide sequence spanning positions 89,726 to 90,301 of pNL1, just upstream from *traD*, is 85% identical to sequences spanning positions 3,990 to 4,576 of a *Sphingomonas* sp. strain RW1 DNA fragment (2). Five ORFs were reported in the 4,576-bp *Sphingomonas* sp. strain RW1 fragment, including the ferredoxin component of dioxin dioxygenase. pNL1 ORF683 corresponds to RW1 ORF4 but has 76 rather than

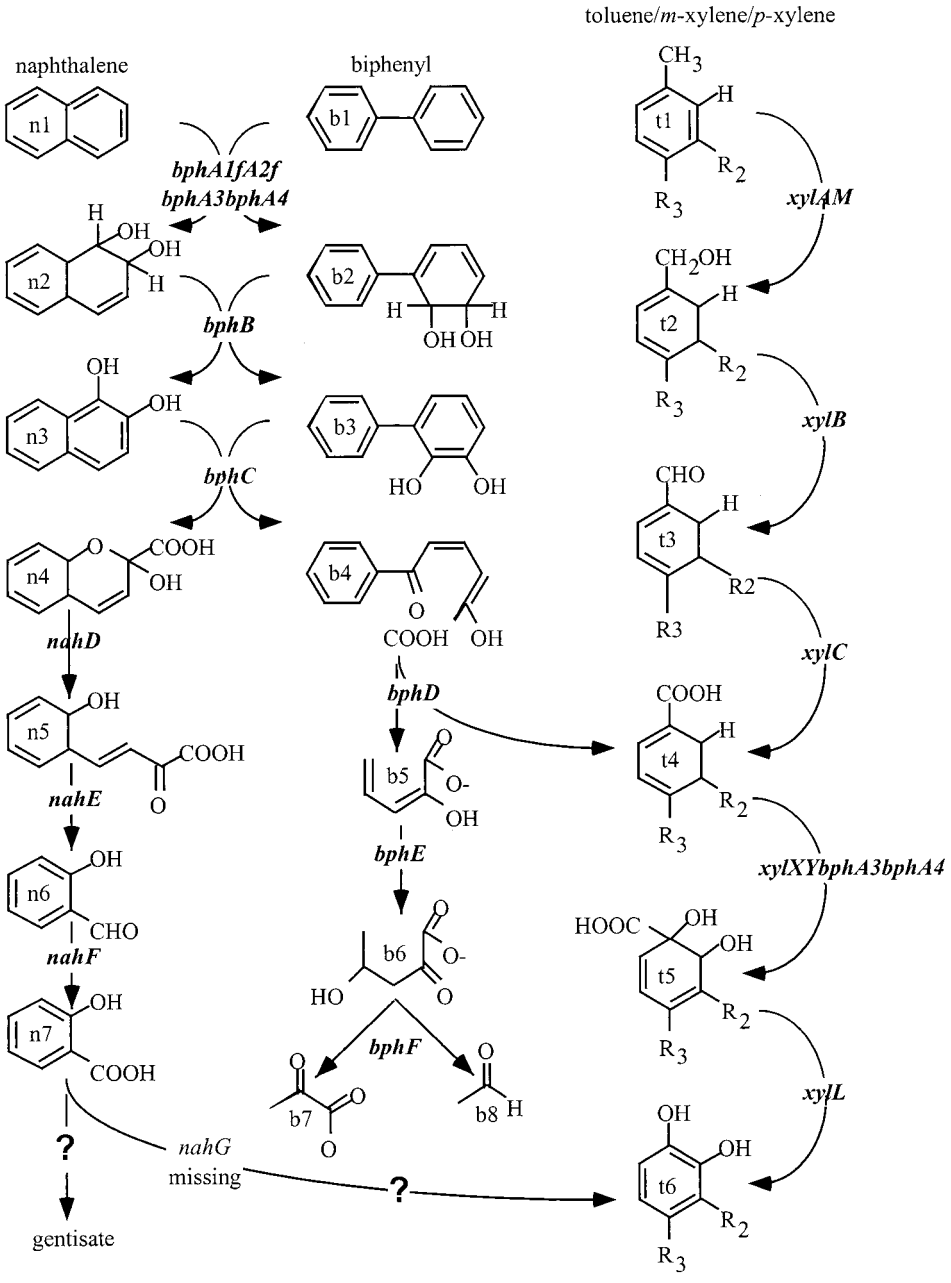


FIG. 4. Hypothetical naphthalene, biphenyl, and toluene degradative pathways encoded on pNL1. Pathway intermediates include naphthalene (n1), *cis*-1,2-dihydroxy-1,2-dihydronaphthalene (n2), 1,2-dihydroxynaphthalene (n3), 2-hydroxy-2-(2-oxo-3,5-cyclohexadienyl)-buta-2,4-dienoate (n4), 2-hydroxychromene-2-carboxylate (n5), salicylaldehyde (n6), salicylate (n7), biphenyl (b1), *cis*-2,3-dihydrodiol (b2), 2,3-dihydroxybiphenyl (b3), 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid (b4), *cis*-2-hydroxy-penta-2,4-dienoate (b5), 4-hydroxy-2-oxovalerate (b6), pyruvate (b7), acetaldehyde (b8), toluene (t1), *m*-methylbenzyl alcohol (t2), *m*-methyl benzaldehyde (t3), benzoate (t4), 1,2-dihydroxycyclohexa-3,5-dienecarboxylate (t5), and catechol (t6). The enzymes predicted to produce these intermediates are dioxygenase (*bphA1fA2fbphA3bphA4*), *cis*-dihydrodiol dehydrogenase (*bphB*), catechol dioxygenase (*bphC*), 2-hydroxychromene-2-carboxylate dehydrogenase (*nahD*), 1,2-dihydroxybenzylpyruvate aldolase (*nahE*), salicylaldehyde dehydrogenase (*nahF*), salicylate hydroxylase (*nahG*), salicylate-5-monooxygenase (*nagG*), 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (*bphD*), 2-oxopent-4-enoate hydratase (*bphE*), 4-hydroxy-2-oxo-valerate aldolase (*bphF*), xylene monooxygenase (*xylAM*), benzyl alcohol dehydrogenase (*xylB*), benzaldehyde dehydrogenase (*xylC*), toluate oxygenase (*xylXYbphA3bphA4*), and dihydroxy cyclohexadiene carboxylate dehydrogenase (*xylL*). $R_1 = R_2 = R_3 = H$ for toluene degradation, $R_1 = R_3 = H$ and $R_2 = CH_3$ for *m*-xylene degradation, and $R_1 = R_2 = H$ and $R_3 = CH_3$ for *p*-xylene degradation.

105 amino acids. However, G+C compensation calculations do not support ORF683 encoding a gene. The location of this region on pNL1 suggests that sequences downstream of the *Sphingomonas* sp. strain RW1 ferredoxin are not related to biodegradation. A preliminary analysis of this region on pNL1 predicts the formation of extensive secondary structure. Among the possible ORFs in the 10-kb gap between the two *tra* clusters, only ORF758 had detectable homology to genes in the public databases.

Nine putative genes were found between *traL* and the *bphA2e*. ORF925 has homology to lytic transglycosylases. Similar proteins encoded by conjugative plasmids have been proposed to facilitate the passage of plasmid DNA through the peptidoglycan layer during conjugation (7, 51). ORF933 and

TABLE 2. Oxygenase activity of cosmid clones encoding *bphC* (cosmid 6) and *xylE* (cosmid 18) with various aromatic substrates

Cosmid	Oxygenase activity (μmol of substrate/min/mg consumed) with:				
	Catechol	3-Methyl catechol	4-Methyl catechol	2,3-Dihydroxy-biphenyl	1,2-Dihydroxy-naphthalene[a]
6	1.03	4.7	1.24	2.73	6.33
18	10.73	14.42	5.77	0.75	0.74

[a] Measured as microliters of O_2 per minute per milligram consumed.

ORF938 are similar to each other, to hypothetical proteins, and to the C-terminal region of the DlpA protein of *L. pneumophila* (8). Interestingly, the entire N-terminal region of DlpA not shared with ORF933 or ORF938 is similar to the entire pNL1 isopropylmalate-tartrate dehydrogenase protein sequence.

Regulatory genes. Six regulatory genes were identified on pNL1. ORF597 is a member of the Ros-MucR family and shares the conserved C_2H_2 zinc finger-like domain, C-X_2-C-X_3-(F,M)-X_5-H-X_4-H associated with this family. ORF569 belongs to the ECF (extracytoplasmic functions) sigma-54 subfamily of regulators (Prosite no. PDOC00814). The ECF sigma factors constitute a diverse group of alternative sigma factors that have been demonstrated to regulate gene expression in response to environmental conditions. ORF758 is a gntR family regulator (Prosite no. PDOC00042) with similarity to regulators that cluster with aromatic catabolic genes in *Acinetobacter* sp. strain ADP1 (unpublished results; GenBank no. AF009672), *P. pseudoalcaligenes* KF707 (unpublished results; GenBank no. 1389649), and *B. cepacia* LB400 (14) and with the regulator for the pNL1 ORF24 homolog, NtaA (94). These similarities suggest that ORF758 may have a role in the regulation of aromatic degradative genes.

ORF154, which is immediately upstream of the cluster of genes encoding the putative pNL1 efflux pump, belongs to the marR family of regulators (Prosite no. PDOC00861). Members of the MarR family respond to environmental signals of the phenolic class such as sodium salicylate (87). Some members of this family, such as EmrR, have been shown to control the expression of efflux pump genes.

BphR is similar to aromatic degradative operon regulators and is a member of the ntcR regulator family (85a). It possesses two of the N-terminal sigma-54 interaction domain signatures (PS00675 and PS00676A) believed to have ATPase activity, as well as a C-terminal HTH lysR regulator family signature (PDOC00043). A region with only a single mismatch to the third sigma-54 interaction domain signature (PS00688) is also present. ORF007 has some similarity to iclR family regulators (Prosite no. PDOC00807) and to *Rhodococcus opacus* CatR (15), which is the regulator for the catechol intradiol cleavage operon. A seventh regulator may be encoded by ORF916, which has weak similarity to hypothetical regulatory proteins.

DISCUSSION

This is the first report of a complete sequence of a conjugative plasmid that encodes pathways for the complete catabolism of aromatic organic compounds. Nearly one-half of the pNL1 DNA encodes genes likely to be involved in either catabolism or transport of aromatic compounds. The remainder of the plasmid appears to encode functions associated with plasmid replication, maintenance, or transfer. A remarkable conservation in both sequence and gene order is evident between pNL1-encoded aromatic catabolic genes and their homologs in other *Sphingomonas* strains. DNA sequences from *S. agrestis* HV3 and *Pseudomonas* sp. strain DJ77 possess high homology to pNL1. A 4,012-bp DNA fragment encoding *cmpFEXC* from *S. agrestis* HV3 plasmid pSKY4 is 90% identical to the pNL1 region encoding *xylF-xylE-cmpX-xylG* (105). DNA fragments from *Pseudomonas* sp. strain DJ77 are 88% identical over a 2,363-bp region encoding homologs (*phnDEF*) to *xylF-xylE-cmpX* (sequence assembled from GenBank entries PSU83881, PSU88298, and PSU83882) and 91% identical over a 1,520-bp region encoding homologs to *xylG-xylJ* (48). The high degree of sequence similarity suggests that either strain DJ77 is a member of the genus *Sphingomonas* or that lateral transfer of genes occurred between *Sphingomonas* and *Pseudomonas* species. The genus *Sphingomonas* is a relatively recent addition to the list of accepted genera (103), and resulted in the renaming of several bacteria (i.e., *P. paucimobilis*, *Flavobacterium capsulatum*, and *Beijerinkia* sp. strain B1).

Thirty-five genes, extending from *bphA2e* to *nahD* are found in the same order and transcriptional direction of the chromo-

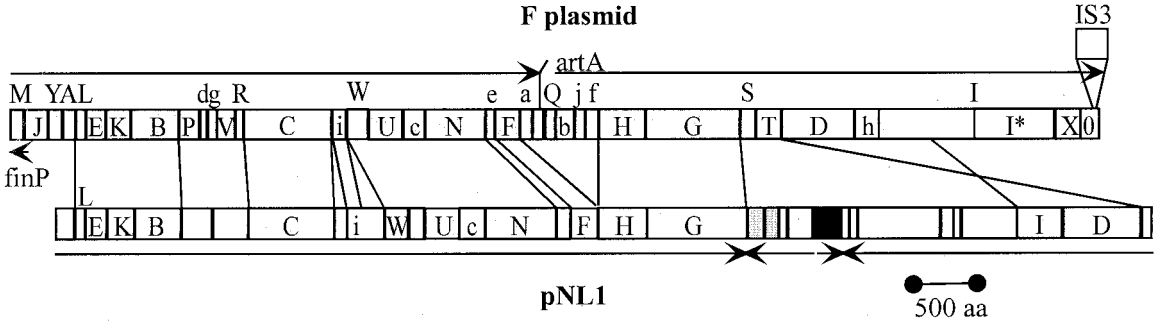


FIG. 5. Comparison of F plasmid and pNL1 transfer regions. Capital letters refer to *tra* genes, and lowercase letters refer to *trb* genes. The direction of gene transcription is indicated by arrows. Lightly shaded boxes depict shared genes, medium-shaded boxes are pNL1 ORFs encoding the ISRm2011-2-like transposase, and the black box is the putative pNL1 regulator encoded by *orf758*. I* refers to the alternative transcript within I. The F plasmid portion of this graphic is from Frost et al. (24).

some of *S. yanoikuyae* B1 (107), an organism that degrades a range of aromatic compounds similar to that of *S. aromaticivorans* F199. The presence of five additional genes (*bphA2c*, *bphA2d*, *cmpX*, *xylT*, and *bphA2e*) not listed in the preliminary report of the *S. yanoikuyae* B1 gene order were confirmed (106). The primary difference between these regions exists at the ends of the 39-kb region. At one end, downstream of the *S. yanoikuyae nahD* gene, is an insertion sequence element. The *bphX* gene at the other end of the *S. yanoikuyae* B1 sequence is a putative membrane protein that is similar to other putative membrane proteins associated with catabolic genes including the *P. fluorescens* IP01 cumene (29) and the *P. putida* F1 toluene (96) degradative gene clusters. Absent from the sequenced *S. yanoikuyae* B1 region were homologs to the putative membrane proteins, ORF1038 and ORF1042, located between *xylM* and *bphA2* in pNL1. Another notable difference between the pNL1 and *S. yanoikuyae* B1 catabolic sequences was the distance between the 3′ ends of *bphB* and *xylA*. The gap between these genes in pNL1 was 60 bp, and in *S. yanoikuyae* B1 it was 555 bp.

Although DNA sequences from *S. yanoikuyae* B1 and *S. aromaticivorans* F199 are similar, significant differences in the catabolic activity of these two strains has been noted. *S. yanoikuyae* B1 is capable of growth on both monocyclic (toluene, *p*-ethyltoluene, 1,2,4-trimethyl benzene, and *m*- and *p*-xylene) and polycyclic aromatics (biphenyl, naphthalene, phenanthrene, and anthracene) (107). *S. aromaticivorans* F199 can also grow on toluene, all xylene isomers, naphthalene, and biphenyl (22), but in our studies growth was not as robust as that of *S. yanoikuyae* B1 on these compounds. Another difference noted was in the ability to express *bphC* and *xylE* in *E. coli* without the need for vector-encoded promoter sequences. *E. coli* containing cosmid clones of *S. yanoikuyae* B1 encoding *xylE* or *bphC* could not be detected by screening for catechol ring cleavage activity (47). In addition, expression of *xylE* from pGEM7Zf(−) was dependent on the vector-encoded *lac* promoter. In contrast, cosmid clones encoding pNL1 *xylE* or *bphC* genes were readily detected in *E. coli* by noting the yellow ring cleavage product that appeared after colonies were sprayed with catechol (86). In sequencing these cosmids, we found that vector-encoded promoters were not required for the expression of ring cleavage activity. Furthermore, the absence of the putative promoter region upstream of *xylF* in cosmid 18 (the present study) suggests that a secondary promoter recognized by the *E. coli* transcriptional machinery is located within the C terminus of *xylF*.

Besides probable differences in gene expression, the presence or absence of specific membrane proteins may also account for the differences in levels of catabolic activity observed in these two *Sphingomonas* strains. Membrane proteins have been implicated in the transport of aromatic compounds across bacterial cell walls. Functions ascribed to these proteins include the export of toxic aromatic compounds out of the cell (2a, 3, 34, 36a, 39) or the uptake of nontoxic aromatics from the environment (9, 67, 78, 100). Aromatic compounds can be toxic due to the adverse effect of accumulation of the hydrophobic compounds in the cellular membrane. Consequently, bacterial growth on aromatic compounds is typically achieved by allowing the substrate to volatilize from a separate phase and diffuse into the medium rather than by direct addition to the growth medium. Some aromatic compounds are not toxic to the cell but are sparingly soluble in water, making them poorly available to bacteria. Therefore, successful growth on aromatic compounds requires both mechanisms for exclusion of toxic aromatics from the cellular interior and uptake of nontoxic aromatic substrates into the cytoplasm where they can be metabolized. A homolog to the functionally uncharacterized membrane protein, BphX, is absent on pNL1 and may play an important role in modulating growth on aromatic compounds which somehow enables *S. yanoikuyae* B1 to grow better in laboratory conditions than *S. aromaticivorans* F199.

The presence of putative efflux pump genes within the aromatic catabolic region of pNL1 suggests that it may function in efflux of aromatic substrates or intermediates associated with pNL1-encoded degradative pathways. Several reports on bacterial efflux pumps associated with solvent (toluene) export and resistance have been published prior to this work (42, 47a, 52a, 79a). The *P. putida* S12 and putative pNL1 efflux pumps are distinct from each other in that they are members of the major facilitator superfamily and the resistance-nodulation-cell division family, respectively. Both families are predicted to utilize proton motive force to drive substrate export (73). It would be interesting to determine if other *Sphingomonas* strains contain similar efflux pump genes.

The export and catabolism of aromatic compounds are two biochemical mechanisms for solvent resistance. Physical mechanisms that lead to increases in cell membrane rigidity are also thought to be important in solvent tolerance (79). This alteration is achieved through desaturation of *cis* fatty acids and subsequent conversion to the *trans* isomer (33, 35, 98). ORF15 and ORFs that cluster with *bphD* are similar to fatty acid biosynthetic pathway enzymes and may prove to function in altering the cellular membrane in response to the presence of aromatic compounds.

The occurrence of multiple oxygenase binding subunit homologs also appears to be characteristic of *Sphingomonas* sp. Although it is tempting to speculate that *S. aromaticivorans* F199 is able to mix and match these oxygenase binding subunits to theoretically form as many as 49 different four-component enzymes, the fact that (i) every large subunit gene is adjacent to a small subunit gene, (ii) members of each large and small subunit pair align best to subunits of the same enzyme, and (iii) all but two of the oxygenase subunit sets are on different transcripts suggests that this does not normally occur. There have been, however, several reports on the construction of hybrid aromatic oxygenases which were in some cases functionally active (72, 89, 93). This suggests that if *S. aromaticivorans* F199 simultaneously expresses several sets of oxygenase binding subunits along with BphA3 and BphA4, novel combinations of oxygenase binding subunits and electron transfer components may occur that result in the formation of hybrid enzymes. Experimentation is needed to test this hypothesis and to measure the ability of hybrid enzymes from this bacterium to oxidize aromatic compounds.

Except for the apparent absence of a *nahG* or *xylZ* homolog, homologs to all genes found in model degradative pathways for biphenyl, *m*-xylene, and naphthalene were found on pNL1. We presume that the *xylZ* ferredoxin and ferredoxin reductase activities are provided by *bphA3* and *bphA4* for the benzoate dioxygenase encoded by pNL1. The existence of similar four-component benzoate dioxygenases have been described in *S. yanoikuyae* B1 (107) and *P. aeruginosa* (81). Since pNL1 conferred the ability to grow with salicylate as the sole carbon and energy source to *Sphingomonas* sp. strain S-88 m260, the genes for catabolism of salicylate are most likely encoded on pNL1. Salicylate can be degraded to gentisate by some microorganisms via catalysis by salicylate-5-monooxygenase (NagG) (25). The BphA1c-BphA2c and BphA1d-BphA2d oxygenase binding subunits cluster most closely with NagG-NagH, to functionally uncharacterized proteins in dinitrotoluene degradative gene clusters, and to the *o*-halobenzoate dioxygenase large subunit, OhbB (Fig. 4). Since *bphA1cA2c* clusters with *nahD*, it

is a prime candidate for an oxygenase catalyzing the degradation of salicylate. Although the gentisate pathway has been described, the sequence for genes in this pathway is limited to two short peptide stretches of 15 and 26 amino acids from gentisate 1,2-dioxygenase (32). No significant homology with these peptides was found among sequences encoded by pNL1. Until more of the gene sequence information is available from the gentisate catabolic pathway, it will not be possible to rule out a gentisate pathway encoded on pNL1.

The putative replication-partitioning region of pNL1 is characteristic of iteron plasmids. The six repeats found between the *orf211* and *orf218 repA* genes are likely iteron sequences to which the RepA binds to regulate plasmid replication. The occurrence of two different repA genes on the same plasmid is unusual and may provide a mechanism for regulating the copy number of pNL1 in response to different environmental stimuli. Bacterial homologs to eukaryotic group II-associated maturases have been previously associated with conjugative elements (60, 66, 104). To our knowledge, this is the first evidence for group II introns on a conjugative plasmid. The unusual distribution of the three putative exons near *matRa* suggest that both maturase-associated splicing and possibly other novel recombinatorial mechanisms are utilized to generate alternate proteins from pNL1.

Overall, the conjugative transfer genes of pNL1 are most similar to those from *E. coli* F plasmid (Fig. 5). The gene order and transcriptional direction of shared genes is the same in pNL1 and F, except that the *traD* and *traI* genes are inverted with respect to the rest of the conjugative genes. F plasmid gene homologs not found on pNL1 include the following: (i) the regulatory genes *traJ*, *traY*, and *finO*; (ii) the pilus synthesis and assembly genes *traA*, *traV*, *traQ*, and *traX*; (iii) the mating-signal gene *traM*, the origin-nicking gene *traY*, and the DNA nicking-unwinding gene *traI** (the C-terminal transcript within TraI); (iv) the surface exclusion genes *traS* and *traT*; and (v) the functionally uncharacterized genes *traP*, *traR*, and *artA* and all *trb* genes save *trbi* and *trbc*. The F plasmid genes are the closest homologs to pNL1 genes except for *traE* (*orf207*), which is most similar to *S. typhimurium* pED208 *traE* (19); *orf851*, whose C-terminal region is homologous to the entire *traF* gene of plasmid RK2 (97); and *orf704*, which is most similar to *E. coli* R388 *trwC* (53).

The presence of multiple ORFs with homology to genes that function in DNA integration and recombination suggests that pNL1 sequences can be rearranged and/or integrated into other genetic elements. The ability to promote site-specific recombination between two closely linked sites on pNL1 is probably provided by the ORF181 invertase. Based on other invertase activities, the predicted outcome of ORF181 activity would be an inversion of a segment of pNL1, most likely in the immediate vicinity of ORF181. Site-specific insertion into a different DNA target is the predicted outcome for phage-type integrases encoded by ORF338, ORF277, and TnpABC. In fact, the localization of the *S. yanoikuyae* B1 aromatic degradative genes on the chromosome may have resulted from such a recombinational event, as early studies with *S. yanoikuyae* B1 (formerly *Beijerinckia* sp. strain B1) suggested that the aromatic degradative genes were plasmid encoded (50).

The sequences of proteins associated with aromatic catabolism in *Sphingomonas* species diverge considerably from those associated with other types of bacteria with similar capabilities. A large proportion of new aromatic catabolic genes sequenced were isolated from *Pseudomonas* species and, not surprisingly, turn out to be very similar to previously characterized genes. This becomes more apparent when doing a cluster analysis of protein families. Identification and characterization of aromatic catabolic genes from phylogenetically distinct bacteria is invaluable in computational predictions of gene functions associated with large sequences generated from microbial genome projects. For example, the characterization and sequencing of the *bphDEF* genes from *Rhodococcus* sp. strain RHA1 (56) allowed us to predict with more reliability that ORFs 1317, 1324, and 1331 encoded *bphDEF*, rather than *xylFJK*, whose products perform very similar functions and whose deduced amino acid sequences are very similar. The same rationale is also true of genes associated with plasmid replication and conjugation. Very few homologs to sequences in the corresponding region of pNL1 were found, making it more difficult to predict functions associated with them. This problem was further compounded by the shuffling of domains among different proteins associated with conjugation (i.e., ORF851 has domains associated with plasmid F TrbI and plasmid RK-2 TraF).

In summary, the analysis of pNL1 sequences suggests that this plasmid confers the ability to utilize and/or tolerate a variety of aromatic compounds to its host, *S. aromaticivorans*. The conferred functions related to aromatic compounds can be divided into three general categories: aromatic catabolism, transport, and cell membrane alteration. The unique codisbursement of genes associated with distinct aromatic catabolic pathways appears to be unique to *Sphingomonas* species and delineates a novel evolutionary path for biodegradative gene organization. The association of membrane proteins with export and/or import of pathway substrates or intermediates is suggested by the sequence annotation. They may provide a means to prevent accumulation of toxic aromatic compounds in the cytoplasm or to sequester them when present at low concentrations in the environment or when they are otherwise poorly available. A protective effect may also be conferred by pNL1-encoded enzymes that change the membrane fatty acid saturation. It should be emphasized that the many of these hypotheses are derived primarily from sequence comparisons and should be considered hypotheses until evaluated experimentally. Work is currently in progress to test some of these predictions.

ACKNOWLEDGMENTS

This work was supported by a Pacific Northwest National Laboratory Initiative in Microbial Biotechnology under Department of Energy contract DE-AC06-76RL0 1830.

We thank Gerben Zylstra for helpful discussions and sharing unpublished data. We also thank Ray Wildung, Blaine Metting, and Ron Walters for their encouragement.

REFERENCES

1. **Armengaud, J., B. Happe, and K. N. Timmis.** 1998. Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. strain RW1: catabolic genes dispersed on the genome. J. Bacteriol. **180:**3954–3966.
2. **Armengaud, J., and K. N. Timmis.** 1997. Molecular characterization of Fdx1, a putidaredoxin-type [2Fe-2S] ferredoxin able to transfer electrons to the dioxin dioxygenase of *Sphingomonas* sp. RW1. Eur. J. Biochem. **247:** 833–842.

2a. **Asako, H., N. Nakajima, K. Kobayashi, M. Kobayashi, and R. Aono.** 1997. Organic solvent tolerance and antibiotic resistance increased by overexpression of *marA* in *Escherichia coli*. Appl. Environ. Microbiol. **63:**1428–1433.

3. **Balkwill, D. L.** 1993. DOE makes subsurface cultures available. ASM News **59:**504–506.
4. **Balkwill, D. L.** 1988. Numbers, diversity, and morphological characteristics of aerobic, chemoheterotrophic bacteria in deep subsurface sediments from a site in South Carolina. Geomicrobiol. J. **7:**33–52.
5. **Balkwill, D. L., G. R. Drake, R. H. Reeves, J. K. Fredrickson, D. C. White, D. B. Ringelberg, D. P. Chandler, M. F. Romine, D. W. Kennedy, and C. M. Spadoni.** 1997. Taxonomic study of aromatic-degrading bacteria from deep-terrestrial-subsurface sediments and description of *Sphingomonas aromaticivorans* sp. nov., *Sphingomonas subterranea* sp. nov., and *Sphingomonas stygia* sp. nov. Int. J. Syst. Bacteriol. **47:**191–201.

6. **Bankier, A. T.** 1993. Generation of random fragments by sonication. Methods Mol. Biol. **23:**47–50.
7. **Bayer, M., R. Eferl, G. Zellnig, K. Teferle, A. Dijkstra, G. Koraimann, and G. Hogenauer.** 1995. Gene 19 of plasmid R1 is required for both efficient conjugative DNA transfer and bacteriophage R17 infection. J. Bacteriol. **177:**4279–4288.
8. **Berger, K. H., J. J. Merriam, and R. R. Isberg.** 1994. Altered intracellular targeting properties associated with mutations in the *Legionella pneumophila dotA* gene. Mol. Microbiol. **14:**809–822.
9. **Collier, L. S., N. N. Nichols, and E. L. Neidle.** 1997. *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. J. Bacteriol. **179:**5943–5946.
10. **Dear, S., and R. Staden.** 1991. A sequence assembly and editing program for efficient management of large projects. DNA Sequence **19:**3907–3911.
11. **Deil, H., C. M. Saint, and P. A. Williams.** 1987. Gene organization of the first catabolic operon of TOL plasmid pWW53: production of indigo by the *xylA* gene product. J. Bacteriol. **169:**764–770.
12. **Ederer, M. M., R. L. Crawford, R. P. Herwig, and C. S. Orser.** 1997. PCP degradation is mediated by closely related strains of the genus *Sphingomonas*. Mol. Ecol. **6:**39–49.
13. **Ensley, B. D., B. J. Ratzkin, T. D. Osslund, M. J. Simon, L. P. Wackett, and D. T. Gibson.** 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science **222:**167–169.
14. **Erickson, B. D., and F. J. Mondello.** 1992. Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated-biphenyl-degrading enzyme in *Pseudomonas* strain LB400. J. Bacteriol. **174:**2903–2912.
15. **Eulberg, D., L. A. Golovleva, and M. Schlomann.** 1997. Characterization of catechol catabolic genes from *Rhodococcus erythropolis* 1CP. J. Bacteriol. **179:**370–381.
16. **Evans, C. A., K. Lewis, and B. E. Rothenberg.** 1989. High efficiency vectors for cosmid microcloning and genomic analysis. Gene **39:**9–20.
17. **Feng, X., L. T. Ou, and A. Ogram.** 1997. Cloning and sequence analysis of a novel insertion element from plasmids harbored by the carbofuran-degrading bacterium, *Sphingomonas* sp. CFO6. Plasmid **37:**169–179.
18. **Figurski, D. H., and D. R. Helinski.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. USA **76:**1648–1652.
19. **Finlay, B. B., L. S. Frost, and W. Paranchych.** 1986. Nucleotide sequence of the *tra* YALE region from IncFV plasmid pED208. J. Bacteriol. **168:**990–998.
20. **Fredrickson, J. K., D. L. Balkwill, G. R. Drake, M. F. Romine, D. B. Ringelberg, and D. C. White.** 1995. Aromatic-degrading *Sphingomonas* isolates from the deep subsurface. Appl. Environ. Microbiol. **61:**1917–1922.
21. **Fredrickson, J. K., D. L. Balkwill, J. M. Zachara, S. W. Li, F. J. Brockman, and M. A. Simmons.** 1991. Physiological diversity and distributions of heterotrophic bacteria in deep cretaceous sediments of the Atlantic coastal plain. Appl. Environ. Microbiol. **57:**402–411.
22. **Fredrickson, J. K., F. J. Brockman, D. J. Workman, S. W. Li, and T. O. Stevens.** 1991. Isolation and characterization of a subsurface bacterium capable of growth on toluene, naphthalene, and other aromatic compounds. Appl. Environ. Microbiol. **57:**796–803.
23. **Freiberg, C., R. Fellay, A. Bairoch, W. J. Broughton, A. Rosenthal, and X. Perret.** 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. Nature **387:**394–401.
24. **Frost, L. S., K. Ippen-Ihler, and R. A. Skurray.** 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. Microbiol. Rev. **58:**162–210.
25. **Fuenmayor, S. L., M. Wild, A. L. Boyes, and P. A. Williams.** 1998. A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J. Bacteriol. **180:**2522–2530.
26. **Furuya, N., and T. Komano.** 1996. Nucleotide sequence and characterization of the *trbABC* region of the IncI1 plasmid R64: existence of the *pnd* gene for plasmid maintenance within the transfer region. J. Bacteriol. **178:**1491–1497.
27. **Gaasterland, T., and C. W. Sensen.** 1996. Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. Biochimie **78:**302–310.
28. **Grothues, D., and B. Tummler.** 1991. New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of *Pseudomonas* species. Mol. Microbiol. **5:**2763–2776.
29. **Habe, H., K. Kasuga, H. Nojiri, H. Yamane, and T. Omori.** 1996. Analysis of cumene (isopropylbenzene) degradation genes from *Pseudomonas fluorescens* IP01. Appl. Environ. Microbiol. **62:**4471–4477.
30. **Happe, B., L. D. Eltis, H. Poth, R. Hedderich, and K. N. Timmis.** 1993. Characterization of 2,2′,3-trihydroxybiphenyl dioxygenase, an extradiol dioxygenase from the dibenzofuran- and dibenzo-*p*-dioxin-degrading bacterium *Sphingomonas* sp. strain RW1. J. Bacteriol. **175:**7313–7320.
31. **Harayama, S., M. Rekik, and K. N. Timmis.** 1986. Genetic analysis of a relaxed substrate specificity aromatic ring dioxygenase, toluate 1,2-dioxygenase, encoded by TOL plasmid pWWO of *Pseudomonas putida*. Mol. Gen. Genet. **202:**226–234.
32. **Harpel, M. R., and J. D. Lipscomb.** 1990. Gentisate 1,2-dioxygenase from pseudomonas. Purification, characterization, and comparison of the enzymes from *Pseudomonas testosteroni* and *Pseudomonas acidovorans*. J. Biol. Chem. **265:**6301–6311.
33. **Heipieper, H. J., R. Diefenbach, and H. Keweloh.** 1992. Conversion of *cis* unsaturated fatty acids to *trans*, a possible mechanism for the protection of phenol-degrading *Pseudomonas putida* P8 from substrate toxicity. Appl. Environ. Microbiol. **58:**1847–1852.
34. **Hofer, B., S. Backhaus, and K. N. Timmis.** 1994. The biphenyl/polychlorinated biphenyl-degradation locus (bph) of *Pseudomonas* sp. LB400 encodes four additional metabolic enzymes. Gene **144:**9–16.
35. **Holtwick, R., F. Meinhardt, and H. Keweloh.** 1997. *cis-trans* Isomerization of unsaturated fatty acids: cloning and sequencing of the *cti* gene from *Pseudomonas putida* P8. Appl. Environ. Microbiol. **63:**4292–4297.
36. **Imai, R., Y. Nagata, M. Fukuda, M. Takagi, and K. Yano.** 1991. Molecular cloning of a *Pseudomonas paucimobilis* gene encoding a 17-kilodalton polypeptide that eliminates HCl molecules from gamma-hexachlorocyclohexane. J. Bacteriol. **173:**6811–6819.

36a. **Isken, S., and J. A. M. de Bont.** 1996. Active efflux of toluene in a solvent-resistant bacterium. J. Bacteriol. **178:**6056–6058.

37. **Jenkins, R. O., and H. Dalton.** 1985. The use of indole as a spectrophotometric assay substrate for toluene dioxygenase. FEMS Microbiol. Lett. **30:**227–231.
38. **Jiang, H., R. E. Parales, N. A. Lynch, and D. T. Gibson.** 1996. Site-directed mutagenesis of conserved amino acids in the alpha subunit of toluene dioxygenase: potential mononuclear non-heme iron coordination sites. J. Bacteriol. **178:**3133–3139.
39. **Kado, C. I., and S.-T. Liu.** 1981. Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. **145:**1365–1373.
40. **Kamoda, S., and Y. Saburi.** 1993. Cloning, expression, and sequence analysis of a lignostilbene-alpha,beta-dioxygenase gene from *Pseudomonas paucimobilis* TMY1009. Biosci. Biotechnol. Biochem. **57:**926–930.
41. **Kato, J., L. Chu, K. Kitano, J. D. DeVault, K. Kimbara, A. M. Chakrabarty, and T. K. Misra.** 1989. Nucleotide sequence of a regulatory region controlling alginate synthesis in *Pseudomonas aeruginosa*: characterization of the *algR2* gene. Gene **84:**31–38.
42. **Kieboom, J., J. J. Dennis, J. A. M. de Bont, and G. J. Zylstra.** 1998. Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. J. Biol. Chem. **273:**85–91.
43. **Kikuchi, Y., Y. Nagata, M. Hinata, K. Kimbara, M. Fukuda, K. Yano, and M. Takagi.** 1994. Identification of *bphA4* gene encoding ferredoxin reductase involved in biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. Appl. Environ. Microbiol. **176:**1689–1694.
44. **Kilpi, S., V. Backstrom, and M. Korhola.** 1980. Degradation of 2-methyl-4-chlorophenoxy acetic acid (MCPA), 2,4-dichlorophenoxy acetic acid (2,4-D), benzoic acid and salicylic acid by *Pseudomonas* sp. HV3. FEMS Microbiol. Lett. **8:**177–182.
45. **Kim, E.** 1996. Ph.D. thesis. Rutgers University, New Brunswick, N.J.
46. **Kim, E., P. J. Aversano, M. F. Romine, R. P. Schneider, and G. J. Zylstra.** 1996. Homology between genes for aromatic hydrocarbon degradation in surface and deep-subsurface *Sphingomonas* strains. Appl. Environ. Microbiol. **62:**1467–1470.
47. **Kim, E., and G. J. Zylstra.** 1995. Molecular and biochemical characterization of two *meta*-cleavage dioxygenases involved in biphenyl and *m*-xylene degradation by *Beijerinckia* sp. strain B1. J. Bacteriol. **177:**3095–3103.

47a. **Kim, K., S. Lee, K. Lee, and D. Lim.** 1998. Isolation and characterization of toluene-sensitive mutants from the toluene-resistant bacterium *Pseudomonas putida* GM73. J. Bacteriol. **180:**3692–3696.

48. **Kim, S., H. J. Shin, Y. Kim, S. J. Kim, and Y. C. Kim.** 1997. Nucleotide sequence of the *Pseudomonas* sp. DJ77 *phnG* gene encoding 2-hydroxymuconic semialdehyde dehydrogenase. Biochem. Biophys. Res. Commun. **240:**41–45.
49. **Kimura, N., A. Nishi, M. Goto, and K. Furukawa.** 1997. Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally. J. Bacteriol. **179:**3936–3943.
50. **Kiyohara, H., M. Sugiyama, F. J. Mondello, D. T. Gibson, and K. Yano.** 1983. Plasmid involvement in the degradation of polycyclic aromatic hydrocarbons by *Beijerinckia* species. Biochem. Biophys. Res. Commun. **111:**939–945.
51. **Koonin, E. V., and R. L. Tatusov.** 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity—application of an iterative approach to database search. J. Mol. Biol. **244:**125–132.
52. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G.**

Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Henaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature **390:**249–256.

52a. **Li, X.-Z., L. Zhang, and K. Poole.** 1998. Role of the multidrug efflux systems of *Pseudomonas aeruginosa* in organic solvent tolerance. J. Bacteriol. **180:** 2987–2991.

53. **Llosa, M., S. Bolland, and F. de la Cruz.** 1994. Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. J. Mol. Biol. **235:**448–464.

54. **Lloyd-Jones, G., and P. C. Lau.** 1997. Glutathione *S*-transferase-encoding gene as a potential probe for environmental bacterial isolates capable of degrading polycyclic aromatic hydrocarbons. Appl. Environ. Microbiol. **63:** 3286–90.

55. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

56. **Masai, E., K. Sugiyama, N. Iwashita, S. Shimizu, J. E. Hauschild, T. Hatta, K. Kimbara, K. Yano, and M. Fukuda.** 1997. The *bphDEF meta*-cleavage pathway genes involved in biphenyl/polychlorinated biphenyl degradation are located on a linear plasmid and separated from the initial *bphACB* genes in *Rhodococcus* sp. strain RHA1. Gene **187:**141–149.

57. **Mercado-Blanco, J., and J. Olivares.** 1994. A protein involved in stabilization of a large non-symbiotic plasmid of *Rhizobium meliloti* shows homology to eukaryotic cytoskeletal proteins and DNA-binding proteins. Gene **139:** 75–79.

58. **Merlin, C., D. Springael, M. Mergeay, and A. Toussaint.** 1997. Organization of the bph gene cluster of transposon Tn*4371*, encoding enzymes for the degradation of biphenyl and 4-chlorobiphenyl compounds. Mol. Gen. Genet. **253:**499–506.

59. **Miele, L., B. Strack, V. Kruft, and E. Lanka.** 1991. Gene organization and nucleotide sequence of the primase region of IncP plasmids RP4 and R751. DNA Sequence **2:**145–162.

60. **Mills, D. J., L. L. McKay, and G. M. Dunny.** 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. J. Bacteriol. **178:**3531–3538.

61. **Missiakas, D., C. Georgopoulos, and S. Raina.** 1994. The *Escherichia coli dsbC* (*xprA*) gene encodes a periplasmic protein involved in disulfide bond formation. EMBO J. **13:**2013–2020.

62. **Moore, E.** 1995. Genetic and serological evidence for the recognition of four pentachlorophenol-degrading bacterial strains as species of the genus *Sphingomonas*. Syst. Appl. Microbiol. **18:**539–548.

63. **Moore, E. R. B., R. M. Wittich, P. Fortnagel, and K. N. Timmis.** 16S ribosomal RNA gene sequence characterization and phylogenetic analysis of a dibenzo-*p*-dioxin degrading isolate within the genus *Sphingomonas*. Lett. Appl. Microbiol., in press.

64. **Mueller, J. G., P. J. Chapman, B. O. Blattmann, and P. H. Pritchard.** 1990. Isolation and characterization of a fluoranthene-utilizing strain of *Pseudomonas paucimobilis*. Appl. Environ. Microbiol. **56:**1079–1086.

65. **Mueller, J. G., R. Devereux, D. L. Santavy, S. E. Lantz, S. G. Willis, and P. H. Pritchard.** 1997. Phylogenetic and physiological comparisons of PAH-degrading bacteria from geographically diverse soils. Antonie Leeuwenhoek **71:**329–343.

66. **Mullant, P., M. Pallen, M. Wilks, J. R. Stephen, and S. Tabaqchali.** 1996. A group II intron in a conjugative transposon from the gram positive bacterium *Clostridium difficile*. Gene **174:**145–150.

67. **Nichols, N. N., and C. S. Harwood.** 1997. PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*. J. Bacteriol. **179:**5056–5061.

68. **Nickel, K., M. J. Suter, and H. P. Kohler.** 1997. Involvement of two alpha-ketoglutarate-dependent dioxygenases in enantioselective degradation of (R)- and (S)-mecoprop by *Sphingomonas herbicidovorans* MH. J. Bacteriol. **179:**6674–6679.

69. **Noda, Y., S. Nishikawa, K.-I. Shiozuka, H. Kadokura, H. Nakajima, K. Yoda, Y. Katayama, N. M. T. Haraguchi, and M. Yamasaki.** 1990. Molecular cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis*. J. Bacteriol. **172:**2704–2709.

70. **Nohynek, L. J., E.-L. Nurmiaho-Lassila, E. L. Suhonen, H.-J. Busse, M. Mohammadi, J. Hantula, F. Rainey, and M. S. Salkinoja-Salonen.** 1996. Description of chlorophenol-degrading *Pseudomonas* sp. strains KF1^T, KF3, and NKF1 as a new species of the genus *Sphingomonas*, *Sphingomonas subarctica* sp. nov. Int. J. Syst. Bacteriol. **46:**1042–1055.

71. **O'Connor, K. E., A. D. Dobson, and S. Hartman.** 1997. Indigo formation by microorganisms expressing styrene monooxygenase activity. Appl. Environ. Microbiol. **63:**4287–4291.

72. **Parales, R. E., M. D. Emig, N. A. Lynch, and D. T. Gibson.** 1998. Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems. J. Bacteriol. **180:**2337–2344.

73. **Paulsen, I. T., M. H. Brown, and R. A. Skurray.** 1996. Proton-dependent multidrug efflux systems. Microbiol. Rev. **60:**575–608.

74. **Paulson, I. T., J. H. Park, P. S. Choi, and M. H. Saier.** 1997. A family of Gram-negative bacterial outer membrane factors that function in the export of protein, carbohydrates, drugs, and heavy metals from Gram-negative bacteria. FEMS Microbiol. Lett. **156:**1–8.

75. **Pflugmacher, U., B. Averhoff, and G. Gottschalk.** 1996. Cloning, sequencing, and expression of isopropylbenzene degradation genes from *Pseudomonas* sp. strain JR1: identification of isopropylbenzene dioxygenase that mediates trichloroethene oxidation. Appl. Environ. Microbiol. **62:**3967–3977.

76. **Philipp, W. J., S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole.** 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proc. Natl. Acad. Sci. USA **7:**3132–3137.

77. **Pollock, T. J., L. Throne, M. Yamazaki, M. J. Mikolajczak, and R. W. Armentrout.** 1994. Mechanism of bacitracin resistance in gram-negative bacteria that synthesize exopolysaccharides. J. Bacteriol. **176:**6229–6237.

78. **Prieto, M. A., and J. L. Garcia.** 1997. Identification of the 4-hydroxyphenylacetate transport gene of *Escherichia coli* W: construction of a highly sensitive cellular biosensor. FEBS Lett. **414:**293–297.

79. **Ramos, J. L., E. Duque, J.-J. Rodríguez-Herva, P. Godoy, A. Haïdour, F. Reyes, and A. Fernández-Barrero.** 1996. Mechanisms for solvent tolerance in bacteria. J. Biol. Chem. **272:**3887–3890.

79a. **Ramos, J. L., D. Estrella, P. Godoy, and A. Segura.** 1998. Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. J. Bacteriol. **180:**3323–3329.

80. **Resnick, S. M., K. Lee, and D. T. Gibson.** 1996. Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. J. Ind. Microbiol. **17:**438–457.

81. **Romanov, V., and R. P. Hausinger.** 1994. *Pseudomonas aeruginosa* 142 uses a three-component ortho-halobenzoate 1,2-dioxygenase for metabolism of 2,4-dichloro- and 2-chlorobenzoate. J. Bacteriol. **176:**3368–3374.

82. **Saint, C. P., and P. Romas.** 1996. 4-Methylphthalate catabolism in *Burkholderia* (*Pseudomonas*) *cepacia* Pc701: a gene encoding a phthalate-specific permease forms part of a novel gene cluster. Microbiology **142:**2407–2418.

83. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

84. **Selbitschka, W., W. Arnold, D. Jording, B. Kosier, N. Toro, and A. Puhler.** 1995. The insertion sequence element ISRm2011-2 belongs to the IS630-Tc1 family of transposable elements and is abundant in *Rhizobium meliloti*. Gene **163:**59–64.

85. **Sharma, S. B., and E. R. Signer.** 1990. Temporal and spatial regulation of the symbiotic genes of *Rhizobium meliloti* in plants revealed by transposon Tn*5-gus*A. Genes Dev. **4:**344–356.

85a. **Shingler, V.** 1996. Signal sensing by σ^{54}-dependent regulators: derepression as a control mechanism. Mol. Microbiol. **19:**409–416.

86. **Stillwell, L. C., S. J. Thurston, R. P. Schnieder, M. F. Romine, J. K. Fredrickson, and J. D. Saffer.** 1995. Physical mapping and characterization of a catabolic plasmid from the deep-subsurface bacterium *Sphingomonas* sp. strain F199. J. Bacteriol. **177:**4537–4539.

87. **Sulavik, M. C., L. F. Gambino, and P. F. Miller.** 1995. The MarR repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. Mol. Med. **1:**436–446.

88. **Sutton, S. D., S. L. Pfaller, J. R. Shann, D. Warshawsky, B. K. Kinkle, and J. R. Vestal.** 1996. Aerobic biodegradation of 4-methylquinoline by a soil bacterium. Appl. Environ. Microbiol. **62:**2910–2914.

89. **Suyama, A., R. Iwakiri, N. Kimura, A. Nishi, K. Nakamura, and K. Furukawa.** 1996. Engineering hybrid pseudomonads capable of utilizing a wide range of aromatic hydrocarbons and of efficient degradation of trichloroethylene. J. Bacteriol. **178:**4039–4046.

90. **Sylvestre, M., M. Sirois, Y. Hurtubise, J. Bergeron, D. Ahmad, F. Shareck, D. Barriault, I. Guillemette, and J. M. Juteau.** 1996. Sequencing of *Comamonas testosteroni* strain B-356-biphenyl/chlorobiphenyl dioxygenase genes: evolutionary relationships among Gram-negative bacterial biphenyl dioxygenases. Gene **174:**195–202.

91. **Taira, K., N. Hayase, N. Arimura, S. Yamashita, T. Miyazaki, and K. Furukawa.** 1988. Cloning and nucleotide sequence of the 2,3-dihydroxybiphenyl dioxygenase gene from the PCB-degrading strain of *Pseudomonas paucimobilis* Q1. Biochemistry **27:**3990–3996.

92. **Taira, K., J. Hirose, S. Hayashida, and K. Furukawa.** 1992. Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes* KF707. J. Biol. Chem. **267:**4844–4853.

93. **Tan, H. M., and C. M. Cheong.** 1994. Substitution of the ISP alpha subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of the enzyme activity. Biochem. Biophys. Res. Commun. **204:**912–917.

94. **Uetz, T., R. Schneider, M. Snozzi, and T. Egli.** 1992. Purification and

characterization of a two-component monooxygenase that hydroxylates nitrilotriacetate from "*Chelatobacter*" strain ATCC 29600. J. Bacteriol. **174:** 1179–1188.
95. **Wang, Y., P. C. K. Lau, and D. K. Button.** 1996. A marine oligobacterium harboring genes known to be part of aromatic hydrocarbon degradation pathways of soil pseudomonads. Appl. Environ. Microbiol. **62:**2169–2173.
96. **Wang, Y., M. Rawlings, D. T. Gibson, D. Labbe, H. Bergeron, R. Brousseau, and P. C. Lau.** 1995. Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in *Pseudomonas putida* F1. Mol. Gen. Genet. **246:**570–579.
97. **Waters, V. L., B. Strack, W. Pansegrau, E. Lanka, and D. G. Guiney.** 1992. Mutational analysis of essential IncP alpha plasmid transfer genes *traF* and *traG* and involvement of *traF* in phage sensitivity. J. Bacteriol. **174:**6666–6673.
98. **Weber, F. J., S. Isken, and J. A. de Bont.** 1994. Cis/trans isomerization of fatty acids as a defence mechanism of *Pseudomonas putida* strains to toxic concentrations of toluene. Microbiology **140**(Pt. 8)**:**2013–2017.
99. **Whited, G. M., and D. T. Gibson.** 1991. Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. J. Bacteriol. **173:**3017–3020.
100. **Williams, P. A., and L. E. Shaw.** 1997. *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous *cis*,*cis*-muconate as the sole carbon source. J. Bacteriol. **179:**5935–5942.
101. **Wong, K. K., L. C. Stillwell, C. A. Dockery, and J. D. Saffer.** 1996. Use of tagged random hexamer amplification (TRHA) to clone and sequence minute quantities of DNA: application to a 180 kb plasmid isolated from *Sphingomonas* F199. Nucleic Acids Res. **24:**3778–3783.
102. **Wood, M. S., C. Lory, and T. G. Lessie.** 1990. Activation of the *lac* genes of Tn*951* by insertion sequences from *Pseudomonas cepacia*. J. Bacteriol. **172:**1719–1724.
103. **Yabuuchi, E., I. Yano, H. Oyaizu, Y. Hashimoto, T. Ezaki, and H. Yamamato.** 1990. Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. Microbiol. Immunol. **34:**99–119.
104. **Yeo, C. C., J. M. Tham, M. W. Yap, and C. L. Poh.** 1997. Group II intron from *Pseudomonas alcaligenes* NCIB9867 (P25X): entrapment in plasmid RP4 and sequence analysis. Microbiology **143**(Pt. 8)**:**2833–2840.
105. **Yrjala, K., L. Paulin, and M. Romantschuk.** 1997. Novel organization of catechol *meta*-pathway genes in *Sphingomonas* sp. HV3 pSKY4 plasmid. FEMS Microbiol. Lett. **154:**403–408.
106. **Zylstra, G. J.** Personal communication.
107. **Zylstra, G. J., and E. Kim.** 1997. Aromatic hydrocarbon degradation by *Sphingomonas yanoikuyae* B1. J. Ind. Microbiol. Biotechnol. **19:**408–414.