Molecular Aspects
of Catabolic Gene Evolution
in Pseudomonads

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH

for the degree of
DOCTOR OF NATURAL SCIENCES

presented by
VLADIMIR SENTCHILO
M.Sc., Belarusian State University, Minsk
born June 25, 1972
citizen of Belarus

Accepted on the recommendation of
Prof. Dr. A. J. B. Zehnder, examiner
Prof. Dr. W.-D. Hardt, co-examiner
Dr. J. R. van der Meer, co-examiner
Prof. Dr. B. Wehrli, chairman

Zürich, 2003
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY</td>
<td>3</td>
</tr>
<tr>
<td>ZUSAMMENFASSUNG</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td>11</td>
</tr>
<tr>
<td>General introduction</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td>41</td>
</tr>
<tr>
<td>Molecular diversity of plasmids bearing genes that encode toluene and</td>
<td></td>
</tr>
<tr>
<td>xylene metabolism in <em>Pseudomonas</em> strains isolated from different</td>
<td></td>
</tr>
<tr>
<td>contaminated sites in Belarus</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 3</td>
<td>73</td>
</tr>
<tr>
<td>Characterization of two alternative promoters for integrase expression</td>
<td></td>
</tr>
<tr>
<td>in the <em>clc</em> genomic island of <em>Pseudomonas</em> sp. strain B13</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td>124</td>
</tr>
<tr>
<td>A case for directed evolution: excision and transfer of the <em>clc</em></td>
<td></td>
</tr>
<tr>
<td>genomic island in <em>Pseudomonas</em> sp. strain B13 is triggered by starvation</td>
<td></td>
</tr>
<tr>
<td>and stimulated by the environmental pollutant, 3-chlorobenzoic acid</td>
<td></td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td>125</td>
</tr>
<tr>
<td>Concluding remarks</td>
<td></td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td>131</td>
</tr>
<tr>
<td>LIST OF PUBLICATIONS</td>
<td>132</td>
</tr>
</tbody>
</table>
Bacterial adaptation towards utilization of xenobiotic compounds as food and energy source largely relies on horizontal gene transfer of mobile genetic elements (MGEs). Hence, to study structural evolution of MGEs and regulation of horizontal gene transfer is of great importance for our understanding of bacterial adaptation. This thesis deals with two classes of MGEs: i) plasmids encoding catabolism of toluene and xylene (TOL plasmids) and ii) conjugative genomic islands, represented by the clc element, which confer its bacterial host the capability to utilize 3-chlorobenzoic acid.

Restriction fragment length polymorphism analysis, DNA-DNA hybridization, physical mapping and partial DNA sequencing were implemented to establish relationships between 12 TOL plasmids (named pSVSs) isolated in Belarus (Chapter 2). The analysis revealed the presence of xyl genes for xylene metabolism on all plasmids. The organization of the xyl genes on the plasmids from strains isolated in Belarus was related to those of other described TOL plasmids, i.e. consisting of upper and meta-pathway operons and their regulatory genes. The xyl operons were detected on three distinctly different types of plasmids (plasmid backbones), suggesting an exchange of complete xyl operons among them. This exchange of ‘gene blocks’ may be the most important mechanism for the appearance of new TOL plasmids. The most prevalent group of plasmids isolated in strains from Belarus (10 out of 12 analyzed in detail) appeared to share strong similarities with the plasmid pWW53, which was isolated from Pseudomonas strains in North Wales (UK). This finding implies that this particular type of TOL plasmid is geographically widespread and is more environmentally and selectively advantageous to its host than other TOL plasmids. Despite the finding that highly similar pSVS plasmids were detected in different bacterial isolates, all of them belonged to the genus Pseudomonas. Even though the metabolic pathway for toluene and xylene degradation can be regarded as evolutionary ancient, the DNA sequences and organization of the xyl genes on the different plasmids gave evidence for further ‘micro-evolutionary’ processes. For instance, clear recombinations were identified in the upper pathway operons of the pSVS plasmids and in the ‘archetypal’ TOL plasmid pWW0. The amount of DNA outside the xyl genes differed by as much as 100 kb between the pSVS plasmids, which is more than half of the total plasmid content, indicating that large plasmid regions are being acquired or lost. We propose, therefore, that plasmid backbones are
constantly exchanging their ‘cargo’ DNA with other replicons with the net effect that suitable genetic functions may end up spontaneously in new recipient cells.

The clc genomic island belongs to a group of integrative conjugative elements whose contribution to horizontal gene transfer and bacterial adaptation is slowly becoming acknowledged (Chapters 3 and 4). Horizontal transfer of the clc island is assumed to depend on the enzymatic activity of a site specific integrase, IntB13, that catalyzes excision of the island from the donor chromosome prior to and its site-specific integration in the new bacterial host chromosome following the conjugation. Expression of intB13 in Pseudomonas sp. strain B13 was studied by using single cell green fluorescent protein (GFP) reporter technology (Chapter 3). Hereto, transcriptional fusions between different intB13 promoter fragments and a promoterless gfp gene were constructed, which were then placed in single copy on to the chromosome of Pseudomonas sp. strain B13. Quantitative GFP fluorescence measurements in individual cells of different B13-derivatives revealed that two alternative integrase promoters exist: P_{circ} and P_{int}. The P_{circ} is a strong constitutive promoter located in the left end of the island and transcribed through the attachment site attP. The corresponding reporter construct was expressed in all cells of Pseudomonas sp. strain B13 and also in E.coli at any growth conditions. We propose that constitutive expression of P_{circ} is needed to provide sufficient amount of integrase to promote integration of the clc island after its transfer to a new bacterial host. Upon integration of the genomic island, the P_{circ} promoter becomes physically disconnected from the coding sequence of the intB13 because the left end (where P_{circ} is located) is displaced from the right end (with the intB13 gene) upon integration. IntB13 expression is now placed under control of the inducible promoter P_{int}. This promoter could be roughly mapped within 232 bp attR region by using reverse-transcription PCR and by successively shortening the promoter fragment from the 5’ end. The characteristics of GFP expression from 5’ deleted P_{int} fragments pointed to possible repressor and activator binding sites in this area. By using transposon mutagenesis we mapped a regulatory region within a 6.4 kb DNA fragment cloned from the left end of the clc island. The DNA sequence of this region and of the transposon insertions pointed to a relatively well conserved area among various other genomic islands, although most of the functions encoded here are unknown. Some homology was found with the parA-parB gene cassette typically found on low-copy-number plasmids. One transposon insertion had a clear down regulating effect on expression of intB13’::gfp, suggesting that this insertion disturbed a gene for an activator protein. The corresponding open reading frame encoded a 175 amino acid protein, but without any significant similarity to functionally characterized proteins in the databases. Transposon
insertions leading to $P_{int}$ overexpression were also found, and mapped in an open reading frame with weak homology to $parB$. However, since other transposon insertions in the same open reading frame did not lead to $intB13'::gfp$ overexpression, we assume at this point that the gene product of this ORF is not a direct repressor for $P_{int}$ expression.

The factors and conditions leading to expression from the integrase promoter $P_{int}$ were further studied in Chapter 4. By using a $P_{int}$-gfp transcription fusion inserted in single copy in Pseudomonas sp. strain B13, we demonstrated its irregular, heterogeneous expression among individual cells of a population. Significantly higher induction levels (i.e., higher GFP fluorescence of individual cells and a higher proportion of cells with detectable GFP fluorescence) were recorded in stationary phase as compared to exponential growth, whereas the presence of the 3-chlorobenzoic acid as opposed to fructose led to a further significant increase. The $intB13$ expression profiles at different physiological states and growth substrates, measured from the $P_{int}$-gfp fusion, correlated well with an increased amount of excised $clc$ island form and increased conjugative transfer of the $clc$ element to $P. putida$ UWC1. Cell density, heat shock, osmotic shock, UV irradiation, or treatment with alcohol had no major influence on integrase expression or led to higher amounts of excised $clc$ element. Hence, we conclude that stationary phase (starvation) conditions and presence of a 3-chlorobenzoate lead to a specific signalling cascade, stimulating the excision and horizontal transfer of the $clc$ genomic island. When interpreted more widely this could mean that ‘directed’ modes of evolution can exist (if one regards horizontal gene transfer as evolutionary event) where growth substrates stimulate the transfer of genetic material for their very metabolism (since the genes for 3-chlorobenzoate degradation are present on the $clc$ element).
ZUSAMMENFASSUNG


für den Toluolabbau auf den belarussischen pSVS Plasmiden deutlich auf eine Rekombination hin, wenn man sie mit denen des TOL-Plasmids pWW0 verglich. Zudem unterschieden sich die pSVS Plasmid-DNAs ausserhalb der Gensequenzen für den Toluolabbau bis zu 100 kb voneinander, was fast die Hälfte des Gesamtplasmids ausmacht. Wir haben aus diesen Ergebnissen die Schlussfolgerungen gezogen, dass Plasmid-DNAs eine Art Rückgratstruktur besitzen, zu der ständig neue Teile herangezogen werden können, vergleichsweise zu einem Schiff, welches regelmässig seine Fracht auswechselt.


Die möglichen Effekte verschiedener Wachstumsbedingungen auf die Expression der Integrase-GFP Fusion vom $P_{\text{int}}$ Promotor wurden im Kapitel 4 weiter untersucht. Dabei stellte sich als erster heraus, dass die Expression des $P_{\text{int}}$ Promotors im Gegensatz zu der des $P_{\text{circ}}$ Promotors fast stochastisch abließ. Dieses Phänomen zeigte sich, indem nur einige Prozente aller individuellen Zellen GFP Fluoreszenz bildeten. Die Veränderungen in der Expression des $P_{\text{int}}$ Promotors wurden deshalb statistisch mittels verteilungsunabhängigen Parametern beschrieben (z.Bsp. 95% Perzentile und 95% Konfidenzintervall). In wiederholten Experimenten konnte festgestellt werden, dass der $P_{\text{int}}$ Promotor in erster Linie in der stationären Phase aktiv wird (wobei immer nur in maximal etwa 10% der Zellen). Während exponentiellem Wachstum ist der Promotor vollständig ausgeschaltet. Das Wachstumssubstrat 3-Chlorbenzoat erwies sich als zweiter zusätzlicher Faktor für die weitere
CHAPTER 1

GENERAL INTRODUCTION

Parts of this chapter were published in Arch. Microbiol. (2001) 175: 79-85 by Jan Roelof van der Meer, Roald R. Ravatn and Vladimir Sentchilo.
**Horizontal gene transfer and bacterial evolution.** The first living cells which emerged on Earth probably about 4 billions years ago were primitive bacteria (61). Since then a tremendous diversity of prokaryotic organisms evolved and occupied all inhabitable niches on our planet. Those niches included other (eukaryotic) living beings, which became hosts for, dependent on or victims of microbes. The tremendous adaptation potential of microbes is the result of plasticity of their genomes and their ability to create genetic variations – the basis for natural selection. Genetic variations can result from the following molecular processes: i) small and local basepair changes within DNA, ii) intragenomic recombination of DNA segments and iii) the acquisition of DNA originating from other (micro)organisms. Acquisition of DNA from other organisms has been termed horizontal or lateral gene transfer and is the main subject of this dissertation.

Curiously, the phenomenon of horizontal gene transfer served as a starting point for the era of molecular biology. It began with the observation of bacterial transformation by O. Griffith in the late 1920s followed by the discovery of the role of DNA in transformation by O. T. Every, C. MacLeod and M. McCarty in 1944. Nowadays, the extent and the importance of horizontal gene transfer for bacterial evolution is becoming clear from comparing whole genome sequences. Comparative genomic studies revealed that bacterial genomes are likely highly dynamic systems steadily receiving and losing DNA. For example, at least 234 lateral transfer events were estimated to have occurred in the *Escherichia coli* genome since this species diverged from the *Salmonella* lineage 100 million years ago (58). This resulted in ca. 18% ‘recent’ DNA present in the current genome permitting *E. coli* to establish itself in otherwise unreachable ecological niches (58). Some serovars of *Salmonella enterica* contain up to 20% DNA not present in the other serovars. A significant part of this ‘variable’ DNA in *Salmonella* contains functions promoting pathogenicity of the organism. At least half of it is thought to have been acquired through horizontal transfer (56). Adaptation to hyperthermophily within moderately thermophilic bacteria was proposed to be another case of genome shaping through DNA acquisition, in this case of genes for thermotolerance from archaean hyperthermophiles (2). Observations of highly related gene functions between different prokaryotic genomes are abundant (43, 69) and will accumulate when more genomes become deciphered. It strongly suggests that horizontal gene transfer processes are and have been extremely important in shaping microbial genomes (42), to the point where one could view a pool of horizontally transferred genes in prokaryotes as one single global organism (42).
In the 20th century the microbial world has been challenged by a massive release of synthetic chemicals with structures and properties to which microorganisms had not been exposed during the course of evolution (so-called xenobiotic chemicals, e.g. chlorinated aromatics). Effective microbial degradation of these compounds therefore required the evolution of new catabolic pathways (44, 90, 112, 113). In many cases such pathways arose by recombination of existing genes from different microorganisms, brought together to one organism by horizontal gene transfer and further fine-tuned by local DNA changes (20, 113, 114). Since various chlorinated compounds are also naturally produced (22), the enzymes which convert them must have been present in bacteria long before the environmental pollution by mankind and could thus be ‘recruited’ to form other metabolic pathways. An example for such recruitment was found in bacteria using chlorobenzenes as sole carbon and energy source. These bacteria carried a specific new arrangement between the genes for chlorocatechol degradation and a set of genes coding for a chlorobenzene dioxygenase and dihydriodiol dehydrogenase (7, 114, 115). Majority of the newly evolved catabolic pathways reside on plasmid DNAs. Plasmids are ideal vehicles for acquiring and collecting DNA from a microbial community. First of all, plasmids are extra-chromosomal self-replicating DNA molecules and therefore possess a high level of flexibility compared to chromosomes by allowing extensive DNA rearrangements without disturbing cell integrity. Secondly, plasmids are intrinsically prone to DNA recombinations due to frequent carriage of IS-elements or transposons. Thirdly and most importantly, large plasmids are self-transmissible and can often mobilize co-resident plasmid replicons. By doing so they transfer, pickup and distribute (potentially) any newly formed catabolic operons throughout microbial communities.

**Plasmids as selfish DNA.** Plasmids may have evolved from primitive replicons resembling current viral RNA ribozymes (49) that have been shown to possess RNA synthesizing and template mediated polymerizing capabilities in the absence of proteins (25, 46). The eventual association of nucleic acids with cells has provided added efficiency and protective advantages for the molecular perpetuation. The promiscuous and selfish nature of plasmids is demonstrated by their ability to engineer their host to better cope with the environment (49), and to maintain their own life style. Plasmids often carry functions advantageous for occupying a specific environment normally inhospitable to the host cell. In this respect plasmid-located operons for catabolism of alternative carbon sources (e.g. xenobiotic compounds) give a growth advantage to their bacterial host. The ‘selfishness’ of plasmids is clear from such functions as copy-number control, a system for multimer resolution, partitioning, post-segregational killing, conjugative transfer and mobilization to
guarantee their own maintenance and distribution. The genes for those functions are organized in the form of inter-regulated and often juxtaposed operons collectively termed the plasmid backbone (105).

In the following I will first deal with plasmids encoding toluene and xylene catabolism as a model to study evolution of catabolic plasmids. The remaining part of introduction is then dedicated to the biology of genomic islands, mobile elements with a clearly different life style but with some mechanistic similarity to self-transmissible plasmids.

TOL plasmids: catabolic pathway and genetic organization. The acronym TOL is a generic name for a group of catabolic plasmids that encode metabolism of toluene, m- and p-xylene via a meta-cleavage pathway (reviewed in (4)). The complete catabolic pathway consists of two sets of reactions (Fig. 1). The first set of reactions (the so called upper pathway) results in conversion of the aromatic compound to its corresponding carboxylic acid, i. e., benzoate, m- or p-toluate. These intermediates are subsequently converted to (methyl-) catechols which are then cleaved in meta position by the catechol 2,3-dioxygenase. A further chain of meta pathway enzymes processes the ring fission product to pyruvate and acetaldehyde (Fig. 1). The genetic organization for and regulation of the toluene and xylene catabolic pathway of plasmid pWW0 (initially named TOL) (121) has been studied in great detail (4). The genes for toluene and xylene degradation (the xyl genes) on pWW0 are clustered in two operons, named the upper and meta pathway operon according to their encoded functions (33-35, 37, 122). The operons are separated by some 14 kb DNA (26) (Fig. 2). Expression of the structural xyl genes on pWW0 is controlled by two positively acting regulatory proteins encoded by the genes xylR and xylS (27, 40), located adjacent to the 3’ end of the meta pathway operon and transcribed divergently (98). XylR stimulates transcription in the presence of pathway substrates from a promoter in front of xylS (Ps) and in front of the upper pathway operon (Pu) (Fig. 2). The XylS protein on its turn stimulates transcription from a promoter in front of the meta cleavage pathway operon (Pm) in the presence of benzoates or upon overexpression of XylS itself. The xylR gene is transcribed more or less constitutively but subject to some repression by XylR itself (62). All the other known TOL plasmids also carry the two operon types with remarkably conserved gene order and with xylS and xylR as regulatory genes, although the number of copies of the operons and regulatory genes on different plasmids may differ (see below).
FIG. 1. Complete metabolic pathway for toluene degradation encoded by TOL plasmids (4). Xylene is funneled through the same pathway. Names of the enzymes are shown in boxes, corresponding structural xyl genes are listed in parentheses. Intermediates (for toluene only) are designated in ovals or numbered.
Transposable elements on TOL plasmids. Several (potentially) mobile DNA elements were discovered on pWW0. First of all, the xyl operons are flanked by two identical insertion sequences in direct orientation and 39 kb apart, named IS\textsubscript{1246} (85). Although their activity has not been demonstrated experimentally, the two copies of IS\textsubscript{1246} have been associated with duplications (94) and deletions of the intervening 39-kb DNA segment (6, 45, 94, 123). The 39-kb region is enclosed within a 56-kb large transposon Tn\textsubscript{4651}. Tn\textsubscript{4651} itself is part of an even larger (70-kb) sized transposon Tn\textsubscript{4653}. Both Tn\textsubscript{4651} and Tn\textsubscript{4653} are functional class II transposons (109-111). The co-integrate resolution system of Tn\textsubscript{4651} is atypical in that its resolvase, TnpS, carries a R-H-R-Y tetrad motif in domains I and II (with proper spacing) which is commonly present only among integrases of the site-specific recombinase family (30). Apart from intramolecular resolution TnpS also catalyzed site-specific intermolecular recombination between two res sites at two different DNA molecules to form one cointegrate (30). On the plasmid pWW53 a 39-kb transposon (Tn\textsubscript{4656}) is found (108). Its transposase, TnpA, and TnpR resolvase are 95% and 87%, respectively, identical and functionally interchangeable with those of Tn\textsubscript{4653} (108).

Both Tn\textsubscript{4651} and Tn\textsubscript{4653} were shown to transpose into various plasmid replicons (15, 45, 59) and to the bacterial chromosome (45, 94). Plasmid free \textit{P. putida} strain MW1000 isolated from the soil carries a Tn\textsubscript{4651}-like TOL transposon in its chromosome (95). Several recombinant plasmids carried, unexpectedly, insertions that differed from two transposons above (59). Likewise recombinants originating from pWW53 and RP4, carried inserts of variable length (4, 51), implying that transposition from pWW0 and pWW53 may sometimes involve DNA segments lying outside the ends of the transposon. The ease of formation of recombinant plasmids between pWW15 and RP4 (52) suggests the presence of transposition functions on pWW15 and, probably, on the related TOL plasmids pWW14 and pWW74 as well (16). New TOL plasmids can also be formed via homologous recombination between their plasmid backbones like was shown for plasmid SAL (for salicylate metabolism) and pWW0 which formed the hybrid plasmid pKF439 (28). The resulting plasmid pKF439 not only acquired both toluene and salicylate catabolism genes, but it also inherited incompatibility determinants of pWW0 (28). A third functional transposable element, the 3,372-bp insertion sequence ISP\textsubscript{pu12}, was just recently described in pWW0 (120). ISP\textsubscript{pu12} occupies the area between the \textit{tnpA} genes of Tn\textsubscript{4651} and Tn\textsubscript{4653} (Fig. 2) and was tentatively assigned to the ISL3 family (116). ISP\textsubscript{pu12} and other closely related elements are widely distributed among gramnegative bacteria and are often found on plasmids and in association
FIG.2. Organization of catabolic xyl genes and associated mobile DNA elements in TOL plasmids. A. Organization of xyl genes on pWW0. Locations of individual xyl genes are marked with arrows. The xyl genes are organized in four transcriptional units: upper pathway operon (filled arrows) (38), meta pathway operon (open arrows) (37) and two regulatory genes, xylS (hatched vertically) and xylR (hatched horizontally). Direction of transcription from corresponding promoters designated Pu, Pm, Ps and Pr is depicted by thin arrows. B. Different arrangements of xyl genes, gene duplications and mobile DNA structures on selected TOL plasmids from different geographical locations. The xyl genes drawn to approximate proportion are designated as in the panel A. Plasmid pWW53 carries two homologous but distinguishable meta pathway operons (assigned meta-I and meta-II) (74), and three not identical homologous xylS genes (depicted S1, S2 and S3) (3). Plasmid pDK1 derived from pWW53 through recombination between xylS1 and xylS3 that was accompanied by inversion of the operon meta-II, deletion of meta-I and production of recombinant xylS1/xylS3 gene (depicted S1/S3) (3). In the plasmid pWW15 upper pathway operon is duplicated, and an incomplete meta pathway operon (assigned “meta”) starting from downstream of xylG is present. In addition pWW15 carries a xylE gene (assigned xylEII) with no homology to the conventional xylE (16, 52). Two overlapping transposons present in pWW0 and one in pWW53 are marked with a solid line under the encompassed DNA region. Functions specified by transposons are displayed as black arrows: tnpA, transposase; tnpS (in Tn4651), integrase-type resolvase; tnpT (in Tn4651), gene stimulating resolution; tnpR (in Tn4653 and Tn4656), resolvase (108-111). Sign ‘(Tn?)’ marks the position of possible transposition functions in pDK1 and pWW15 which both recombine with plasmid RP4 (52, 92). The insertion sequence ISPpu12 (120) and two copies of IS1246 (85) on pWW0 are each marked by connected and oppositely faced open triangles.
with catabolic genes (116). The element is highly active, can form multiple insertions and can even provoke large-scale genome rearrangements (116, 120).

**Gene duplications in TOL plasmids.** Plasmid pWW0 carries the \( xyl \) genes in single copy only. A similar set of \( xyl \) operons as on pWW0 is present on pWW102, although in a different arrangement and divergent in DNA sequence (1). Many TOL plasmids, however, carry duplications of \( xyl \) genes. For example, a second highly homologous yet distinguishable copy of the meta pathway operon is present on pWW53 (74) and on three related TOL plasmids pWW5, pWW74 and pWW88 (16). Plasmids pWW15, pWW14, pWW20 and pWW74 (16, 52, 124) carry duplications of both \( xyl \) operons, with one of the meta pathway operons being incomplete (68, 119). Two non-identical copies of the \( xylS \) regulatory gene are found on pDK1 (3, 92) and three on pWW53 (3). Gene duplications may have some advantage for the organism by increasing gene dosage and thus enzyme expression levels (see (54) and references within). Secondly, variant alleles may produce isoenzymes or heteromultimers with improved catalytic performance. Thirdly, regulation of different duplicated transcription units may confer more flexibility in responding to changing environmental conditions. This point is supported by the observation that transcription of either of the two \( xylS \) homologues on pWW53 was dependent on the growth substrate (29). Fourthly, the presence of non-identical gene copies may favor new genetic variations by recombination or mutation.

**Diversity and evolution of TOL plasmids.** Our knowledge about the evolution of the \( xyl \) genes is mainly based on pWW0. From several studies it was concluded that (i) the meta pathway operon originated as a fusion product of two independently evolved gene blocks, \( xylXYZL \) and \( xylTEGFJQKIH \) (36); (ii) the upper and meta pathway operons evolved separately (32); and (iii) the meta pathway operon, in the strict sense (i.e., \( xylT \) through \( xylH \)), probably resulted from a recombination event itself (8, 36). It was suggested that the transcription regulator \( xylR \) and the corresponding operator-promoter region upstream of the upper pathway operon were recruited from another regulatory system (32). It is possible that the meta pathway operon of TOL plasmid pWW102 (1) is a mosaic structure as well. This assumption is based on the finding that the \( xylQ \) gene of pWW102 is more homologous to \( nahO \) of the naphthalene catabolism plasmid pWW60-22 (79) than to corresponding sequences of pWW0 and pDK1 (1). Much less information is available on the backbone structures of TOL plasmids. It was established that pWW0 belongs to the IncP-9 plasmid incompatibility group (5) and is a broad-host-range plasmid (9). For example, pWW0 (or a plasmid not distinguishable by restriction analysis) was present in five different *Pseudomonas*
isolates from Wales (23). An *Alcaligenes eutrophus* strain 345 (39) was found to carry a TOL plasmid slightly smaller than pWW0, but with very similar restriction profiles. Other studied TOL plasmids were shown to belong to different and even unknown incompatibility groups. They also differed considerably in size (ranged between 100 and more than 250 kbp) and restriction enzyme profiles from each other and from pWW0 (Chapter 2). Furthermore, not all TOL plasmids are transmissible, indicating that *xyl* genes are carried on different plasmid backbones. Whether distribution of the whole *xyl* pathway to different plasmids is associated with the transposons Tn\textsuperscript{4651}, Tn\textsuperscript{4653} and Tn\textsuperscript{4656} can not be firmly concluded due to lack of detailed knowledge about the plasmid backbones and the DNA regions flanking the *xyl* genes on the different TOL plasmids. As already mentioned above a number of TOL plasmids carries duplicated *xyl* gene clusters. Whether these duplications were generated via recombination with other replicons or resulted from one ancient duplication event and kept afterwards remains an open question. Duplications are potentially unstable and prone to deletions of intervening DNA via homologous recombination, which has often been observed in organisms maintained in the lab during cultivation on benzoate (6, 52, 55, 77, 92). However, double copies of *xyl* operons seem to persist in natural isolates pointing at some selective advantage (Chapter 2).

In conclusion, *xyl* genes on TOL plasmids are associated with insertion sequences and transposons and are prone to intra-replicon rearrangements and inter-replicon exchange. This implies that potentially almost any plasmid (and chromosome as well) may sequester *xyl* genes in the form of a transposon and selectively propagate in the environment providing substrates for the *xyl*-encoded pathway. If so an environmental pool of *xyl* genes may exist, which steadily undergoes further divergent evolution. Owing to the variety of existing plasmid replicons *xyl* genes can be maintained in different bacterial hosts and be exchanged with those of other TOL plasmids. To find out whether these assumptions were correct we performed in this thesis a comparative analysis of TOL plasmids isolated from bacteria prevailing in environments polluted with oil products (Chapter 2).

The *clc* element of *Pseudomonas sp.* strain B13. In an almost classical experiment, chlorobenzene-degrading organism had been constructed in the laboratory by mimicking natural gene transfer (73, 87). For this purpose, *Pseudomonas sp.* strain B13, a bacterium that metabolizes 3-chlorobenzoate (and carries the genes for chlorocatechol degradation) was mated with *Pseudomonas putida* F1, which degrades toluene. Bacteria using chlorobenzene as sole carbon and energy source were obtained, which were of *P. putida* F1 origin, indicating
that they had acquired a piece of DNA from strain B13 (73). The same mating method was used to transfer the genes for chlorocatechol degradation from strain B13 to various other recipients (86). Until recently, it had always been suspected that the genes were transferred from strain B13 by means of a conjugative plasmid, pB13 (17). This plasmid had been described once, however, independent isolation remained unsuccessful (117).

More recently, upon repetition of this mating experiment, it has been found out that the transferred element was not a classical plasmid. In fact, in P. putida F1 the transferred DNA behaved rather strangely and formed large, unstable duplications, which were integrated into the chromosome at two locations (82). The transferred DNA was at that point named the clc element (for its first described characteristics: chlorocatechol degradation). Subsequent analysis of strain B13 itself showed that the element was present also in the chromosome, in two separate locations (82). This indicated that the element acted as an integrative plasmid or perhaps as a conjugative transposon. The size of the transferred element was determined from large chromosomal fragment mapping and from cosmid libraries as 105 kb (82). Restriction site mapping indicated an almost perfect match to the published physical map of the described plasmid pB13, which suggested that the clc element could be present in two forms, as free plasmid and integrated. Since the element was present in two locations on the chromosome, it was possible to compare both integration sites and determine the ends of the element. This indicated that the right end was contained near or in a gene for glycine tRNA (tRNA$^{Gly}$) whereas the left end was formed by an 18-bp sequence repeated from the right end (Fig. 3). About 250 bp downstream from the right end, a large open reading frame was found coding for a phage-type integrase (83). This suggested that the clc element was not just an integrated plasmid, but employed an integrase to achieve site-specific integration. Activity of the integrase could be shown in Escherichia coli by using a two-plasmid system: one plasmid with attP upstream the integrase (like in the circular form of the element) and another plasmid with the reconstructed attB site (containing tRNA$^{Gly}$) (83). When attR was placed upstream the integrase gene (like in the integrated situation) no recombination was detected.

**Bacteriophage-related genetic elements.** From the analyses described above, it became clear that the clc element is a bacteriophage-related element. Such elements are a loosely defined group of DNA elements, including some bacteriophages, prophages, insertional plasmids, conjugative transposons and ‘genomic islands’ (most of which are known as pathogenicity islands, PAIs). The distinctions and definitions given in the literature sometimes become rather diffuse and artificial. In fact, it seems as if a spectrum of different mobilization and transfer mechanisms is being utilized. All these elements are related by their
means of site-specific recombination, involving integrase enzyme and (in some cases) excisionase. Integration sites are often formed by the 3'-end of tRNA genes (mostly for genomic islands and some prophages) but may also be other sequences (like for conjugative transposons). Genomic islands usually are present in only 1 copy per cell, whereas conjugative transposons may retain more copies (and more target sites) per cell (13). Recently, the more general name ‘integrative and conjugative elements’ (ICE) was suggested.

FIG. 3. Mechanism of sequence-specific integration of the clc element. A). Under normal conditions in Pseudomonas sp. strain B13, the element is integrated in two tRNA\textsuperscript{Gly} locations. The most distal 18-bp of the tRNA\textsuperscript{Gly} are repeated at the left end (attL) (shown as white and black triangles). The integrase gene (intB13) is located near the right end (attR). In the circular intermediate the left and right ends are joined (attP), generating a strong promoter for the integrase (arrow). B. With a certain frequency, the clc element transfers to a new host and again integrates into the chromosomal target sequence (tRNA\textsuperscript{Gly}). The 18-bp 3'-end of the tRNA\textsuperscript{Gly} gene is displaced to the left end, but the original tRNA\textsuperscript{Gly} sequence is restored by an identical 18-bp sequence of the clc element (open triangle).
to cover this group (13). Apart from this, however, the remaining DNA of the element may encode totally different functions, varying from phage structural proteins to plant symbiotic factors or virulence determinants. For example, most genomes contain integrated prophages, although it is not always clear if they can be induced. Satellite phages, like P4 (78) or the *Staphylococcus aureus* pathogenicity island SaPI (72), require another phage (e.g., P2 for P4) for head assembly and transduction. Other bacteriophages (P1 or P4) have a plasmid origin of replication (11, 75); however, their mode of transfer is by transduction. Insertional plasmids, as found in Actinomycetes (97) seem to replicate autonomously in most hosts, but can also integrate into the chromosome. The IncJ element R391 from *Proteus retgeri* behaves as an insertional plasmid. In *E. coli*, it is normally integrated at one unique site in the chromosome. However, the element has functional plasmid replicative and incompatibility systems, and plasmid forms can be isolated from *recA* strains in which the chromosomal site is occupied by a related element (10, 76). Insertional plasmids use conjugation rather than transduction as means of transfer. Conjugative transposons (e.g., CTn916) mostly do not replicate independently, but form a circular intermediate upon excision, from which conjugal transfer is initiated (91). PAIs mostly occur as unstable regions of the chromosome and carry virulence genes (31). Some of them still have active integration functions and may even behave as phages such as the 40-kb VPI pathogenicity island of *Vibrio cholerae* (50). Other PAIs contain an integrase and a replication origin, like the 12-kb *vap* region of the Gram-negative anaerobic pathogen *Dichelobacter nodosus* (18). Some PAIs, like the high pathogenicity island (HPI) of *Yersinia pseudotuberculosis* still carry a phage-type integrase, but do not show evidence for functions involved in self-transmissability or replication (12). And in some PAIs, even the integrase function has become lost (31).

On the basis of similarities among integrase enzymes, five large groups are recognized (67). These consist of i) bacteriophage-type integrases, ii) resolvases, iii) transposases, iv) excisionase/integrases, and v) invertases. All these proteins essentially maintain the active-site residues (the HXXRX_{28-35}Y signature), although the overall amino acid sequence similarities may be insignificantly low (67). The use of the term ‘integrase’, therefore, causes considerable confusion, because it does not directly infer its specific mode of action. For example, resolvases will promote site-specific recombination between two copies of a specifically recognized direct repeat sequence on the same DNA. On the other hand, bacteriophage-type integrases promote cutting-out (excision) and subsequent integration into a new target DNA (not necessarily on the same DNA molecule). The integrase from the *clc* element belongs to the bacteriophage-type group (83) and is related to what is called the
phage P4 subfamily of integrases. These integrases are not only site-specific but also sequence-specific, and use tRNA structural genes as integration sites (Fig. 4). Interestingly, various other elements employ P4-type integrases (Fig. 4) although there is no direct conservation in the targeted tRNA genes. P4-type integrases have been found in prophages themselves; in PAIs (such as the vap element of D. nodosus (18), or the HPI of Y. pseudotuberculosis (12), and in a ‘symbiosis island’ in Mesorhizobium loti. The symbiosis island encompasses approximately 500 kb and contains the genes necessary for nodule formation, nitrogen fixation and synthesis of three vitamins (101, 102).

The percentages of amino acid identity and similarity of the IntB13 integrase are moderately low (32/55%) with respect to the P4-type integrases mentioned above (Fig. 4) and are limited to an N-terminal region of approximately 420 amino acids. In addition, the integrase of the clc element is about 250 amino acids longer than most other P4-type integrases. This was long puzzling, since only two other sequences (i. e., those in P. aeruginosa and Pseudomonas strain B4) were available which showed significant similarity to this C-terminal extension, (Fig. 4). It might be that these are actually part of a larger open reading frame, which has not yet been determined. Recently, several new integrase sequences of the IntB13 type, i. e. those with extended C-terminus were deposited in GenBank. An almost identical gene as the intB13 integrase is found in the unfinished genome sequence of Burkholderia fungorum (at present only unaligned contigs are available) (Fig. 4). Other integrase genes highly similar to intB13 were detected in the complete genomes of the plant pathogens Xylella fastidiosa (93) and Xanthomonas campestris (21). In X. campestris even two distinct intB13 homologues occur (Fig. 4). Interestingly, a gene encoding only the C-domain of both IntB13 and X. fastidiosa integrase is also present elsewhere on the X. fastidiosa genome (Fig. 4). This ORF might have been part of an ancient intact IntB13-type integrase gene, or rather a precursor which when recombined with an ancestral P4-type integrase formed the longer IntB13-type integrase. Integrase genes highly related to IntB13 were also found on two genomic islands, i. e. PAGI-2 and PAGI-3, of Pseudomonas aeruginosa (57) and in Ralstonia metallidurans (unfinished genome project), although in the latter the intB13-like gene seemed disrupted (Fig. 4). Strangely enough, the homology in most cases, except between strain B13 and B. fungorum, stops immediately downstream of the integrase gene (Fig. 4).
FIG. 4 (on the left). Comparison of the gene organizations of integrases related to IntB13 of *Pseudomonas* sp. strain B13. Arrows indicate sizes and direction of the major open reading frames. Gene names or gene assignments are indicated directly below the open reading frames. At the far left, the names of bacteria in which the integrative elements were originally found and the GenBank accession numbers are listed. Note that seven integrases most closely related to the intB13 are considerably longer than other P4-type integrases and, in fact, are longer than most other known integrases (67). Shading in grey corresponds to those parts carrying similarity to P4-type integrases. Shading in black corresponds to genes (or parts of genes) which exclusively carry similarities to the C-terminal end of the intB13. Numbers below the integrase ORFs indicate % identity, % similarity, number of gaps, and length of the overlap for the deduced amino acid sequences, respectively.

Screening of the *X. fastidiosa* sequence downstream of the putative integrase gene revealed that possible repetitions of the 18-bp recognition sequence of the *tRNA^Gly* gene are found at three locations; at 66.6, 71.6 and 365 kb distance downstream of the *tRNA^Gly* gene (GenBank entry AE004839). The sequence at a distance of 71.6 kb had again strong similarity with the left end of the *clc* element and may, therefore, mark the end of an unrecognized genomic island in *X. fastidiosa*. The left end sequence of the *clc* element of B13 was also relatively well conserved to sequences in *B. fungorum* and *R. metallidurans*, and to the left ends of the two *P. aeruginosa* pathogenicity islands, but was lacking in *X. campestris* (Chapter 3). Most of the open reading frames in the left end of the *X. fastidiosa* ‘genomic island’ region and on the *P. aeruginosa* pathogenicity islands encode hypothetical proteins (57, 93). However, some ORFs indicate weak relationships to conjugative transfer capabilities and phage-type genes. Sequencing of the *clc* element is currently ongoing and it is too early to present a detailed comparison between the *clc* element and the *X. fastidiosa* and *P. aeruginosa* genomic islands, but several strongly conserved regions are found in all four genomic islands (*X. fastidiosa*, B13 and two in *P. aeruginosa*) (57). This suggests that genome islands exemplified by the intB13 type integrase have conserved core structures, which perhaps encode transfer, maintainance or regulatory functions. Perhaps *X. campestris* once also carried an intact genome island, but lost a major part of it. Apart from *Pseudomonas* sp. strain B13 two other carriers of the intB13 type integrase also possess genes for chlorocatechol catabolism (*clc*): *B. fungorum* and *Ralstonia* sp. strain JS705. *Ralstonia* sp. strain JS705 carries a complete mobilizable *clc* type element (65). *B. fungorum* may also carry a complete *clc* element. This assumption is based on BLAST analysis of a roughly 25 kb DNA region covering left and right ends of the *clc* element. For example, the genes for chlorocatechol catabolism of B13 display 99-100% nucleotide identity to those in *B. fungorum*. 
**Integration mechanism.** There is too little information to draw conclusions on all details of integration, excision and transfer of the *clc* element. However, based on experiments and similarities to other systems, we can propose the following mechanisms. In all hosts analyzed so far, the *clc* element is present in the integrated form in one, two or more copies (Fig. 3) (82, 83). With a certain low frequency, the element excises, which results in a circular intermediate. Whether the circularized molecule is capable of replicating remains to be determined. The fact that the *clc* element previously had been characterized as a plasmid pB13 (17), might indicate that in some hosts replication may occur. Also the observation of multiple tandem copies integrated in the chromosome of *P. putida* F1 might point to replication (82). Analysis of the *attP* region on the circular form indicated that ends of the *clc* element are connected here. No indication was found for a heterologous hybrid as formed in the circular intermediate of conjugative transposons (91). The *attP* sequence carries an identical 18-bp sequence as the *attB* site of the target, which seems to be the actual site of recombination. If the *clc* element works similar as conjugative transposons, we should assume that the element excises precisely from its original location and that the cuts introduced in this area are ligated. In fact, PCR analysis of cultures of strain B13 jim1-gfp (Chapter 3) suggested that the integration site was restored upon excision, leaving no target duplications behind. There may well be an equilibrium state between the excised and the integrated form. With a certain frequency, however, the element transfers to another recipient (84) which in analogy to conjugative transposons may involve single-stranded transfer and subsequent replication. The net result would be that the donor cell does not lose the transferred element. The transfer mechanism is basically unknown; however requirement for cell-to-cell contact (V. Sentchilo, unpublished) points to conjugation rather than transduction. In the new host, the element integrates very effectively, since connecting both ends in the circular form creates a strong promoter for the integrase [Chapter 3 and 4, (83)]. A prerequisite for the target site is a \( tRNA^{Gly} \) gene and, probably, some sort of hairpin structure immediately downstream of the target. This can be concluded from comparison of insertion sites in different hosts in which the *clc* element has integrated. For example, *P. aeruginosa* has three copies for \( tRNA^{Gly} \), however, only the one with the downstream hairpin is targeted. *Pseudomonas putida* has four copies for \( tRNA^{Gly} \), but only two of them (carrying the hairpin structure) are used as integration sites. *X. fastidiosa* (93) has two copies of the \( tRNA^{Gly} \) gene, but only near the one with the hairpin structure, the IntB13-like integrase is present (Fig. 5).
FIG. 5. Comparison of the integration sites for the clc element in different recipient bacteria and for the (putative) X. fastidiosa element. For reasons unknown, the element integrates only into tRNA<sup>Gly</sup> genes, which carry a downstream sequence of strong diad symmetry. The black arrow points to the actual site of insertion.

**Regulation of integrase expression.** For most ICEs (prophages, genomic islands and conjugative transposons) the integrase plays an essential (although not exclusive) role in excision. This was concluded from insertions or mutations in genes for integrase and the resulting lack of formation of the excised product. Apart from the integrase itself, excisionase factors – or recombination directionality factors (RDF) - are often required. It is thus justified to consider the integrase expression as a ‘first step’ in activation of integrated elements. For most ICEs, still relatively little is known on regulation of integrase expression, although regulatory cascades have been discovered governing lysogenic/lytic switches (for bacteriophages) or for excision and conjugative transfer (conjugative transposons). For example, the *Vibrio cholerae* pathogenicity island VPI is a self-transmissible entity, spreading as a filamentous phage termed VPIφ (50). Phage particles of VPIφ consist of monomers of the protein TcpA which is also a building block for the type-IV pili required for pathogenicity and can serve as a receptor for the toxin producing phage CTXφ. Expression of TcpA occurs simultaneously with that of the cholera-toxin and is dependent on environmental conditions.
like temperature, osmolarity, pH and composition of the growth medium (64, 96). It has not become clear if phage particle production is occurring at the same time as expression of tcpA and the toxin genes. Phage particles as well as double stranded phage chromosomes of VPIφ were isolated from overnight cultures of V. cholerae (50) suggesting that there is no specific induction needed for phage production, but that phages ‘leak’ out of the cell continuously.

Regulation of excision and phage production works differently for the satellite bacteriophage P4. P4 and also the closely related retron phage φR73 are defective phages, which are dependent on helper phages, like P2, to produce virions (41, 78). Depending on the presence of P2 in the host cell, the life of P4 will end up in lysogenic state, multicopy self-replicating plasmid (1% of cases) or lytic cycle (11). The presence of P2 favors the lytic cycle. The regulatory cascade behind the lysogenic/lytic switch in P4 is very complex, involving various life cycle dependent promoters and transcription factors both from P4 itself and from P2 (11, 47, 48, 89). Expression of the P4 integrase is under control of two promoters. One of them apparently belongs to the σ70 type with well conserved –10 and –35 elements (78), and expression from this promoter is high immediately after P4 infection. In a later stage expression from this promoter is repressed 1000-fold due to binding of the IntP4 integrase to it (78). The DNA sequence of another integrase promoter, about 100 bp downstream, displays similarity to the consensus of the so-called late promoters of phage P2. Therefore it was suggested that this integrase promoter may be regulated by P2 factors (78). The specific conditions for excision are not known, although coinfection with P2 increases excision frequency (89). Excision requires a second protein, Vis (or Orf88), which binds to the attP region and overlaps with the –35 region of the int promoter.

In the cryptic P4-like prophage CP4-57 of E. coli a gene alpA was reported to regulate transcription of the integrase gene slpA (107). Induction of the integrase gene led to 100% excision and loss of the prophage (53). This was based on supplementing alpA on a high-copy-number plasmid to strains containing the prophage. Without supplementing alpA the slpA integrase gene was transcribed at a very low basal level. Some sequence similarities can be detected in the upstream regions of the integrase gene in P4, φR73 and CP4-57, suggesting DNA binding proteins to bind here (107). The original promoter of alpA in the prophage state itself was apparently silent. This agrees with stable maintenance of the prophage in E. coli K-12. AlpA has only 28% amino acid identity to Vis (Orf88) of P4, but is assumed to have a similar function (53). Interestingly, proteins homologous to AlpA/Vis have also been found in HPI of Y. pestis and Y. pseudotuberculosis (81).
All excision mechanisms result in a circular double-stranded intermediate which can then transfer to a new recipient. In the case of phage transmission, a capsid is build around the excised DNA. For the *clc* element, phage transduction and transformation seem less likely as mechanism for cell to cell transfer, since no *clc* transfer was detected when cell-cell contact was avoided. Therefore, conjugation seems the more likely mode of transfer for the *clc* element. Although both conjugative plasmids and integrative conjugative elements are now known to share mechanistically similar transfer steps and display conserved transfer genes, genomic islands of the *clc* type carry no genes homologous to known *tra* or *mob* determinants [(13, 57, 106) and references within].

Another issue which has not been well studied is signalling and regulation of conjugative transfer, despite the impressive amount of accumulated structural and mechanistic information (104). Probably the best examples for regulation and signalling in conjugative transfer consist of the opine-regulated Ti plasmids in *Agrobacterium* [e. g. (70, 71)] and the pheromone-stimulated *Enterococcus* plasmids. *Enterococcus* elaborate a plasmid-collection mechanism that allows recipient cells to stimulate transfer by the donor by secreting a polypeptide pheromone. Pheromone-induced plasmids confer either antibiotic resistance determinants or virulence factors (24, 103, 125). Regulation of transfer of two conjugative transposons is also quite well studied. Transfer of the conjugative transposon CTn916 is stimulated 5 to 10-fold when cells have been grown in the presence of tetracycline. Tetracycline causes antitermination at several terminator structures, the result of which is i) a mRNA covering the complete CTn916 (circular excised) element, ii) transcription of the *tra* genes, and iii) activation of two regulatory genes orf7 and orf8, which produce a second long transcript covering the *tra* genes (14). Strangely enough, excision itself, which is dependent on the CTn916 integrase, is not stimulated by tetracycline. The transfer of the conjugative transposons CTnDOT and CTnERL from *Bacteroides* is also stimulated when cells are grown in the presence of tetracycline, but even up to 10,000-fold. The regulatory mechanisms for CTnDOT are different from CTn916, though. Tetracycline triggers transcription of an operon containing *tetQ* (tetracycline resistance gene) and two regulators *rteA* and *rteB*. RteA and RteB in turn activate transcription of *rteC* (100). RteC positively affects expression of the topoisomerase-like protein Exc and the latter stimulates both excision and expression of the *tra* genes (19, 118). Tetracycline does not influence transcription of the integrase gene itself.

Observations made with *Pseudomonas* sp. strain B13 indicated that transfer of the *clc* element might be stimulated by the compound whose very metabolism it encodes, i. e. 3-chlorobenzoic acid (CBA). Two observations made in our lab pointed to this: (i) increased
frequency of the clc island transfer upon addition of low concentrations of CBA to the lab-
scale activated sludge microcosm and (ii) appearance of the circular form of clc element in
cell cultures grown on CBA as opposed to complete medium. In this thesis an attempt was
made to decipher the molecular mechanisms of this stimulation (Chapters 3 and 4).

**Is the clc element a curiosity or the tip of an iceberg?** Until very recently clc element
was the only known site-specific integrative self-transmissible element encoding ‘novel’
catabolic function. However, more data are accumulating that mobile elements like the clc
genomic island have an important part to play in catabolic gene transfer and pathway
evolution and transfer of pathogenicity determinants. We are aware of at least two bacteria
carrying an element very strongly related to the clc element. One of them is already
mentioned *B. fungorum* that perhaps contains an identical clc element. The second bacterium,
*Ralstonia* sp. strain JS705 (65) was isolated from contaminated groundwater in the USA,
grows on chlorobenzene and 3-chlorobenzoate, and carries the clc genes for chlorocatechol
degradation (115). Yet other examples for a self-transmissible chromosomal elements with
catabolic genes exist. One of these comes from studies on the biphenyl-degrading organism *P.
putida* KF715 (66). *P. putida* KF715 carries a 90-kb conjugative element which was first
discovered as an unstable chromosomal region. Deletions were found to arise spontaneously,
which involved loss of either or both the bph and sal genes (for biphenyl and salicylate
degradation, respectively) or large-scale chromosomal rearrangements. Mating with strain
KF715 resulted in the transfer of both *bph* and *sal* genes from the original strain to *P. putida*
AC30, and from this strain to *P. putida* KT2440. In both recipient organisms, the element
could be targeted to the chromosome. Unfortunately, no information is available on the
characterization of the genes for integration, mobilization, on the target site, or mode of action
of this element. It might be that the bph-sal element is related to the other biphenyl mobile
element Tn4371 (see below) or to the partial integrase sequences of *P. aeruginosa* and
*Pseudomonas* sp. strain B4 (Fig. 4), which are also located near naphthalene and biphenyl
degradation genes.

The presence of a phage-type integrase was demonstrated for the transposon structure
Tn4371 in *Ralstonia eutropha* (63, 99). The transposon is 55 kb in size and carries a 13-kb
region with the bph genes for biphenyl degradation. The right end contains a 40-kb second
transposon, tn-bph. At the left end, an integrase gene is found. This integrase, however, has
very little overall sequence similarity to IntB13, but rather belongs to the phage P22
subfamily. Tn4371 inserts site-specifically into a TTTTTCAT-sequence, which is present on
RP4 in two sites, and in five sites in the chromosome of *R. eutropha*. The transposon does not
form cointegrates, indicating that it has a cut-and-paste mode of transposition, rather than replicative mode. A circular intermediate of Tn4371 could be detected by PCR. Precise excision was observed from the kanamycin resistance gene within RP4, which was restored exactly. Although the right end of Tn4371 contains two ORFs with similarity to Ti-membrane transfer proteins, no independent conjugation of Tn4371 has yet been detected (63). Possibly Tn4371 is a derivative of a previously self-transmissible integrative element that lost the ability to transfer autonomously but ‘learned’ how to travel on conjugative IncP-1 plasmids. Two other biphenyl catabolism-encoding integrative elements related to Tn4371 and tentatively assigned Tn4372 and Tn4373 were described by Springael et al. (99). They differ substantially from each other but also displayed common features. They have similar size, carry highly similar bph genes, coexist with and are mobilized by similar pSS50-like plasmids. However, Tn4372 and Tn4373 did not hybridize with the left-end (i.e. integrase containing) part of Tn4371 and differ in their insertion sites on RP4. This indicated a different nature of the integration functions in all three elements. Therefore, the authors proposed that Tn4371, Tn4372 and Tn4373 originated via combination of different functional modules for catabolism of biphenyl and for integration. Tn4372 and Tn4373 were not able to transfer on their own.

In three other bacteria with degradative pathways have there been descriptions of possible integrase functions, to the best of our knowledge. Two of these (i.e., in Pseudomonas pavonacea and Mycobacterium sp. G1) occurred in gene regions for a dehalogenase involved in haloalkane degradation (80). The other occurred on the catabolic plasmid pNL1 of Sphingomonas aromaticivorans (88). However, all of these integrases belong to the ‘resolvase’ class, rather than the bacteriophage-class.

In conclusion, Pseudomonas sp. strain B13 carries a transferable element, which is unique for bacteria that degrade aromatic or aliphatic pollutants. This element, which employs a phage P4-like integrase, is related to a family of elements that are widespread among bacteria (Fig. 4). It may be representative for a new subgroup of elements with C-extended integrases formed by the putative ‘genomic islands’ in B. fungorum, X. fastidiosa, X. campestris, in Ralstonia sp. strain JS705 and in P. aeruginosa (Fig. 4).

Scope and outline of this thesis. In this thesis representatives of two classes of mobile DNA elements were studied. By investigating the molecular diversity of the TOL
plasmids we aimed to track major routes and mechanisms of their evolution. Regulation of horizontal gene transfer was studied using the example of the \textit{clc} genomic island.

Chapter 2 describes the molecular analysis of the TOL plasmids isolated from different contaminated sites in Belarus. By doing restriction fragment length polymorphism analysis, by constructing complete restriction maps, by DNA-DNA hybridization, and by sequencing several \textit{xyl} gene regions we have established the relationships between catabolic genes and plasmid backbones on different TOL plasmids. We propose here that gene exchange of various scales, acquisition of \textit{xyl} operons by non-catabolic plasmids, and horizontal spread of TOL plasmids throughout microbial community have played and continue to play major role in the evolution of \textit{xyl} pathway.

In the Chapter 3 two differentially expressed promoters of the integrase were characterized governing integrase expression in either integrated or the excised circular form of the \textit{clc} genomic island. Integrase promoter-\textit{gfp} fusions were used to monitor expression of the integrase gene in individual cells. Positive and negative regulation of the integrase transcription was disclosed and corresponding regulatory functions were mapped near the left end of the \textit{clc} island. This DNA region appeared to be conserved in several other genomic islands in various microorganisms including plant and human pathogens, thus shading light on evolutionary relationships of the genomic island in \textit{Pseudomonas} sp. strain B13.

Regulation of the integrase, excision and horizontal transfer of the \textit{clc} island in response to changing environmental conditions were addressed in Chapter 4. Expression of the integrase in \textit{Pseudomonas} sp. strain B13 was monitored using transcriptional fusion of the integrase promoter (as in the integrated island) to \textit{gfp} reporter gene. It was shown that the integrase is specifically induced in small subpopulation of cells during stationary phase conditions and is not affected by such stress factors as UV irradiation, heat, high osmolarity or ethanol. Remarkably, induction of the integrase, excision and horizontal transfer of the \textit{clc} element were positively influenced by growth on 3-chlorobenzoic acid, i. e. the substrate whose metabolism is encoded by the island itself, that may point to the directed evolution. Selfish nature of the \textit{clc} island became apparent from the observation that upon the onset of unfavorable conditions (e. g. starvation) \textit{clc} island ‘escapes’ to another bacterial host via conjugation.
REFERENCES


Molecular diversity of plasmids bearing genes that encode toluene and xylene metabolism in *Pseudomonas* strains isolated from different contaminated sites in Belarus

ABSTRACT

Twenty different *Pseudomonas* strains utilizing *m*-toluate were isolated from oil-contaminated soil samples near Minsk, Belarus. Seventeen of these isolates carried plasmids ranging in size from 78 to about 200 kb (assigned pSVS plasmids) and encoding the *meta* cleavage pathway for toluene metabolism. Most plasmids were conjugative, but of unknown incompatibility groups, except for one, which belonged to IncP9 group. The organization of the genes for toluene catabolism was determined by restriction analysis and hybridization with *xyl* gene probes of pWW0. The majority of the plasmids carried *xyl*-type genes highly homologous to those of pWW53 and organized in similar manner (Gallegos, M. T., P. A. Williams, and J. L. Ramos. 1998, J. Bacteriol. 179: 5024-5029), with two distinguishable *meta* pathway operons, one upper pathway operon, and three *xylS*-homologous regions. All of these plasmids also possessed large areas of homologous DNA outside the catabolic genes, suggesting a common ancestry. Two other pSVS plasmids carried only one *meta* pathway operon, one upper pathway operon and one copy each of *xylS* and *xylR*. The plasmid backbones of these two differed greatly from the others. Whereas the parts of the plasmids carrying *xyl* genes were mostly conserved between plasmids of each group, the noncatabolic part had undergone intensive DNA rearrangements. DNA sequencing data of specific regions near and within the *xylTE* and *xylA* genes of the pSVS-plasmids confirmed the strong homologies to *xyl* genes of pWW53 and pWW0. However several recombinations were discovered within the upper pathway operons of the pSVS plasmids and pWW0. The main genetic mechanisms which are thought to have resulted in the present day configuration of the *xyl* operons are discussed in light of the diversity analysis carried out on the pSVS plasmids.
INTRODUCTION

The pathway for the catabolism of toluene, \textit{m}- and \textit{p}-xylenes via aromatic ring \textit{meta} cleavage is often encoded on plasmids, referred to as TOL-plasmids (1, 4 [and references within], 8, 13, 26, 28, 29, 31, 37, 54, 55, 57, 58). TOL plasmids have mostly been found in representatives of the genus \textit{Pseudomonas}, with one exception for \textit{Alcaligenes eutrophus} strain 345 (pRA1000) (21), isolated from a variety of geographical locations. The plasmids differ in size, restriction pattern, compatibility, and conjugation ability, although the \textit{xyl} genes for toluene/xyylene catabolism are mostly organized in the \textit{meta} pathway operon (\textit{xylXYZLTEGRKQJH}) and an upper pathway operon [\textit{xyl(UW)CMAB(N)}]. Expression of the \textit{xyl} genes is under control of two regulatory genes, \textit{xylS} and \textit{xylR} (15, 41). There is a high level of DNA homology among \textit{xyl} genes on different TOL plasmids, and the gene order within either \textit{meta} or upper pathway operons is invariable. However, the relative positions of both the operons and regulatory genes vary (1, 12, 15, 47). In some TOL plasmids, more than one copy of the regulatory genes and/or \textit{xyl} operons has been found (3, 8, 37, 38, 53).

A number of TOL plasmids have been shown to recombine DNA with other plasmids such as antibiotic resistance (R) plasmids (23, 26, 37, 47) and a plasmid for salicylate degradation (14), and with the host chromosome (25, 33, 48). Furthermore, on pWW0 the \textit{xyl} genes are located within a 56-kb transposon (Tn4651), which itself is part of a larger 73-kb transposon (Tn4653) (51). The \textit{xyl} genes on plasmids pWW53 are also part of a transposon, Tn4656, but for pDK1 and pWW15 the precise mechanism for \textit{xyl} gene translocation is not yet established.

Knowledge about the evolution of the \textit{xyl} genes is mainly based on pWW0. From several studies, it was concluded that (i) the \textit{meta} pathway operon originated as a fusion product of two independently evolved gene blocks, \textit{xylXYZL} and \textit{xylTEGRKQJH} (19); (ii) the upper and \textit{meta} pathway operons evolved separately (17); and (iii) the \textit{meta} pathway operon in strict sense (i.e., \textit{xylT} through \textit{H}), probably resulted from a recombination itself (6, 19). However, due to limited sequence information on DNA regions flanking the \textit{xyl} genes in different TOL plasmids, no solid conclusions can be drawn about the evolution of \textit{xyl} genes and the mechanisms leading to their distribution and variation on different plasmid replicons.

The objective of this study was to compare a variety of TOL plasmids in order to establish trends in their evolution and the underlying evolutionary mechanisms. For this purpose, we isolated a number of \textit{m}-toluate degrading strains from oil-contaminated sites in
Belarus, located the genes for toluene catabolism on plasmids present in those strains, and compared their organization and DNA sequence. Relationships between plasmid backbones and between \textit{xyl} gene clusters on the different TOL plasmids were established and evolutionary mechanisms proposed.
MATERIALS AND METHODS

Media and growth conditions. Luria-Bertani (LB) medium and M9 minimal medium (MM) (46) were routinely used. Sodium benzoate, m-toluate or salicylate was added to MM to a final concentration of 5.0 mM; phenol was used at 2.5 mM and glucose was used at 20 mM. Toluene and xylene were supplied through the vapour phase to MM agar plates incubated in gas-tight glass jars. When required, the medium was supplemented with ampicillin at 100 µg/ml for Escherichia coli and 500 µg/ml for Pseudomonas; kanamycin at 50 µg/ml; gentamycin at 50 µg/ml; rifampin at 50 µg/ml; streptomycin at 50 µg/ml; tetracycline at 30 µg/ml; nalidixic acid at 50 µg/ml; 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), at 0.004% w/v; or L-histidine at 40 µg/ml. Strains of Pseudomonas were grown at 30°C, and those of E. coli were grown at 37°C.

Bacterial strains and plasmids. Bacterial strains and plasmids used in this study are listed in Tables 1 and 2. Twenty Pseudomonas sp. strains described in this study were isolated from 13 soil samples collected from distant locations (from 3 to 100 km apart) near Minsk (Belarus). All but two sites were near oil transportation routes and have a history of major contamination by oil products (Table 1). Samples were handled under conditions preventing bacterial cross-contamination. Initial enrichments for toluene- and xylene-degrading microorganisms were made in 250 ml Erlenmeyer flasks containing 30 ml of minimal medium with m-toluate, which were inoculated with 10 g of soil sample and incubated for 24 hours at 27°C and 150 rpm on a rotary shaker. The resulting cultures were serially diluted and plated on MM agar plates with m-toluate (TMA). Colonies visible after 48 hours were purified by streaking on TMA and were tested for their ability to grow on benzoate, toluene, m-xylene, phenol or salicylate in liquid medium and on agar plates. Growth was judged by comparison with that on MM without carbon source. Activity of catechol 2,3-dioxygenase (C2,3O) was tested by using the catechol spray test (12). Pure cultures were stored in LB agar under paraffin oil at 4°C and frozen in m-toluate-minimal medium with 20% glycerol at –80°C. Identification of the isolates was performed on the basis of morphological and cultural characteristics according to Bergey's Manual of Systematic Bacteriology (1984) (30) and by sequencing of a ribosomal 16S rDNA fragments amplified by the polymerase chain reaction (PCR).
TABLE 1. *Pseudomonas* strains isolated in the present study

<table>
<thead>
<tr>
<th><em>Pseudomonas</em> strain</th>
<th>Relevant genotype or characteristics</th>
<th>Soil sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV1</td>
<td>Ap', Te', Na', Cm', Th', mXln', mTol', Ben', C2,3O³; carries a 2.3-kb cryptic plasmid</td>
<td>Layer of solidified oil on the surface of a long-distance oil transportation tank; railway station Minsk-South</td>
</tr>
<tr>
<td>SV2</td>
<td>Ap', Te', Cm', Th', mXln', mTol', Ben', C2,3O³</td>
<td>Same as SV1</td>
</tr>
<tr>
<td>SV3</td>
<td>Ap', Te', Cm', Th', mXln', mTol', Ben', C2,3O³; carries ~190-kb plasmid pSVS15</td>
<td>Sandy soil contaminated with heating oil; southwestern part of Minsk</td>
</tr>
<tr>
<td>SV4</td>
<td>Te', Na', Cm', Th', mXln', mTol', Ben', C2,3O³; carries ~200-kb plasmid pSVS4</td>
<td>Same as SV3</td>
</tr>
<tr>
<td>SV5</td>
<td>Te', Th', mXln', mTol', Ben', C2,3O³; carries ~150-kb plasmid pSVS12 and a 2.9-kb cryptic plasmid</td>
<td>Railway bed soil contaminated with mixture of lubricants, diesel fuel, and phenol; town of Smorgon, 100 km to North-West from Minsk</td>
</tr>
<tr>
<td>SV6</td>
<td>Te', Th', mXln', mTol', Ben', C2,3O³, Sal'; carries 140-kb plasmid pSVS16 and cryptic 10-kb and 80-kb plasmids.</td>
<td>Grass bed bulk soil, contaminated with diesel fuel; central part of the city</td>
</tr>
<tr>
<td>SV9</td>
<td>Ap', Te', Cm', Th', mXln', mTol', Ben', C2,3O³; carries 107-kb plasmid pSVS9</td>
<td>Railway bed soil contaminated with mixture of lubricants, diesel fuel and phenol; central part of Minsk</td>
</tr>
<tr>
<td>SV10</td>
<td>Ap', Th', mXln', mTol', Ben', C2,3O³; carries 105-kb plasmid pSVS10</td>
<td>Same as SV9</td>
</tr>
<tr>
<td>SV11</td>
<td>Te', Th', mXln', mTol', Ben', C2,3O³; carries 130-kb plasmid pSVS11</td>
<td>Railway bed soil contaminated with mixture of lubricants, diesel fuel, and phenol; northwestern part of Minsk</td>
</tr>
<tr>
<td>SV12</td>
<td>Ap', Th', mXln', mTol', Ben', C2,3O³; carries 195-kb plasmid pSVS12</td>
<td>Park soil with no history of contamination known; northeastern part of the city of Minsk</td>
</tr>
<tr>
<td>SV13</td>
<td>Na', Th', mXln', mTol', Ben', C2,3O³; carries 78-kb plasmid pSVS13</td>
<td>Same as SV12</td>
</tr>
<tr>
<td>SV15</td>
<td>Cm', Hg', Th', mXln', mTol', Ben', C2,3O³; carries 90-kb plasmid pSVS15</td>
<td>Piece of rubber from a storage of used car tires; central part of Minsk</td>
</tr>
<tr>
<td>SV16</td>
<td>Ap', Cm', Hg', Th', mXln', mTol', Ben', C2,3O³; carries 123-kb plasmid pSVS16</td>
<td>Sandy soil contaminated with diesel oil; building area in northeastern part of Minsk</td>
</tr>
<tr>
<td>SV17</td>
<td>Ap', Te', Th', mXln', mTol', Ben', C2,3O³; carries ~200-kb plasmid pSVS17</td>
<td>Sandy washout from the highway, contaminated with a mixture of oil products; central part of Minsk</td>
</tr>
<tr>
<td>SV19</td>
<td>Ap', Te', Cm', Hg', Th', mXln', mTol', Ben', C2,3O³, Sal'; carries 107-kb plasmid pSVS19</td>
<td>Shore line sediments contaminated with a mixture of oil products; river Swislach, central part of Minsk</td>
</tr>
<tr>
<td>SV20</td>
<td>Ap', Hg', Th', mXln', mTol', Ben', C2,3O³; carries plasmid pSVS20 (&gt;200 kb)</td>
<td>Same as SV19</td>
</tr>
<tr>
<td>SV22</td>
<td>Ap', Km', Te', Th', mXln', mTol', Ben', C2,3O³; carries 126-kb plasmid pSVS22</td>
<td>Sandy soil contaminated with a mixture of industrial lubricants and engine oil; local landfill 10 km from the southwestern border</td>
</tr>
<tr>
<td>SV23</td>
<td>Ap', Th', mXln', mTol', Ben', C2,3O³; carries 130-kb plasmid pSVS23</td>
<td>Same as SV22</td>
</tr>
<tr>
<td>SV24</td>
<td>Th', mXln', mTol', Ben', C2,3O³; carries 105-kb plasmid pSVS24</td>
<td>Same as SV22</td>
</tr>
<tr>
<td>SV25</td>
<td>Cm', Th', mXln', mTol', Ben', C2,3O³; carries a 2-kb cryptic plasmid</td>
<td>Road cover material, local landfill 10 km from the southwestern border of Minsk</td>
</tr>
</tbody>
</table>

* Ap', Te', Km', Na', Cm', and Hg', resistance to antibiotics ampicillin, tetracycline, kanamycin, nalidixic acid, chloramphenicol and HgCl₂, respectively, in concentrations stated in Materials and Methods. Th', mXln', mTol', Ben', Sal', ability to use toluene, m-xylene, m-toluolate, benzoate, and salicylate, respectively as sole carbon and energy source, respectively; C2,3O³, positive for catechol 2,3-dioxygenase activity.
**Plasmid curing experiments.** Plasmid free and plasmid deletion derivatives of the toluene degrading strains were isolated during repeated cultivation on benzoate (31, 56). The catechol spray test was used to distinguish wild type from any mutant cells lacking C2,3O activity. Wild-type colonies turned bright yellow due to production of 2-hydroxymuconic semialdehyde by C2,3O, whereas the colonies formed by plasmid-free cells or by those carrying deletions in their plasmid DNAs, resulting in loss of the gene for C2,3O, remained white. Colonies formed by mutants with reduced level of C2,3O activity turned pale yellow. The sensitivity of detection of mutants on plates was approximately 1 mutant among 10,000 colonies. Mutants were purified by streaking on LB agar, and the relevant catabolic phenotypes were tested by plating onto solid minimal media supplemented with the appropriate carbon sources. In a few cases when growth on benzoate did not produce any plasmid mutants, mitomycin C was added to the liquid benzoate-minimal medium at concentrations of 1.25, 2.5, 5.0, or 7.5 µg/ml.

**Filter mating experiments.** Donor and recipient strains were grown overnight at 30°C in 5 ml LB medium to a cell density of about 1x10⁹ cells/ml. Cells from 1 ml of each culture were harvested by centrifugation, resuspended each in 1 ml of fresh LB, mixed and incubated for 3 h at 27°C without shaking. The resulting conjugation mixture was harvested by a 5 min spin at 3,000 rpm in a microcentrifuge (Hereaus AG, Zurich, Switzerland) and transferred by gentle pipetting onto 0.45-µm-pore-size cellulose nitrate filters (Sartorius AG, Göttingen, Germany). The filters were placed on LB agar plates, which were incubated for 24 h at room temperature. Cells were then washed from the filter with a 0.9% NaCl solution, serially diluted and plated onto MM or LB agar plates supplemented with the appropriate C-source, antibiotics or amino acids. Separate cultures of the donor and recipient strain were treated and plated in the same way to provide a negative control. Colonies of transconjugants were purified by streaking on selective agar plates and tested for the presence of relevant phenotypic markers. Plasmid DNA was isolated and analyzed by agarose gel electrophoresis directly or after digestion with the appropriate endonucleases.

**Plasmid incompatibility test.** Plasmid RP1, belonging to the plasmid incompatibility group IncP1 (22), and pMG18, belonging to the group IncP9 (25) were introduced into SV strains (Table 1) through the filter mating procedure (see above). Transconjugants were selected on glucose-minimal agar plates with the appropriate antibiotics plus 0.05 mM of m-toluate to induce C2,3O. Colonies grown for 48 h were checked by the catechol-spray test for
### TABLE 2. Bacterial strains and plasmids

<table>
<thead>
<tr>
<th>Bacterial strains and plasmids</th>
<th>Relevant characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> DH5α</td>
<td>F supE44 (φ80dlacZAM15) Δlac(lacZYA-argF)U169 hsdR17 recA1 endA1 gyrA96 (Nal⁺) thi-1 relA deoR</td>
<td>Gibco BRL, Life Technologies</td>
</tr>
<tr>
<td><strong>E. coli</strong> ED8654</td>
<td>supE supF hsdR metB lacY gal trpR; host for propagation of pWW53-3001 and pWW53-3002</td>
<td>P. A. Williams</td>
</tr>
<tr>
<td><strong>E. coli</strong> LE392 (RP1)</td>
<td>Plasmid of IncP1 incompatibility group; Km², Tc²</td>
<td>S. Harayama</td>
</tr>
<tr>
<td><em>P. putida</em> mt-2</td>
<td>Tln⁺, mXln⁺, mTol⁺, Ben⁺, C2,3O⁺; host of pWW0</td>
<td>P. A. Williams</td>
</tr>
<tr>
<td><em>P. putida</em> MT53</td>
<td>Tln⁺, mXln⁺, mTol⁺, Ben⁺, C2,3O⁺; host of pWW53</td>
<td>P. A. Williams</td>
</tr>
<tr>
<td><em>P. putida</em> AC13</td>
<td>His⁺, Rif⁺, Sm²; recipient strain for TOL plasmids</td>
<td>A. Boronin</td>
</tr>
<tr>
<td><em>P. putida</em> AC34 (pMG18)</td>
<td>Plasmid of IncP9 incompatibility group; Sm², Gm², Km², Hg²</td>
<td>G. Jacoby</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWW53-3001</td>
<td>HindIII 17.5 kb fragment (HA) of pWW53 cloned in pKT230; contains <em>meta</em> pathway operon II</td>
<td>P. A. Williams</td>
</tr>
<tr>
<td>pWW53-3002</td>
<td>HindIII 15.6 kb fragment (HB) of pWW53 cloned in pKT230; contains <em>meta</em> pathway operon I</td>
<td>P. A. Williams</td>
</tr>
<tr>
<td>pUC18</td>
<td>Ap⁺, ColE1 replicon; general cloning vector</td>
<td>Promega Corp.</td>
</tr>
<tr>
<td>pUC28</td>
<td>Ap⁺, ColE1 replicon; general cloning vector</td>
<td>Promega Corp.</td>
</tr>
<tr>
<td>pGEM-T Easy</td>
<td>Ap⁺, ColE1, linearized pGEM-5Zf(+) with single 3’-thymidine overhangs to facilitate cloning of PCR products</td>
<td>Promega Corp.</td>
</tr>
<tr>
<td>pPL392</td>
<td>16.4-kb EcoRI fragment of pWW0; contains complete <em>meta</em> cleavage pathway operon and <em>xylS</em> cloned into pBR322</td>
<td>(18)</td>
</tr>
<tr>
<td>pCVS22</td>
<td>6.3-kb EcoRI fragment of pSVS11 cloned into pUC18, contains <em>xylW</em> (3’ end) and <em>CMABN</em> (5’ end)</td>
<td>This study</td>
</tr>
<tr>
<td>pCVS31</td>
<td>5.3-kb HpaI fragment of pSVS10 cloned into pUC28; contains genes <em>xylC</em> (3’ end) and <em>MABN</em> (5’ end)</td>
<td>This study</td>
</tr>
<tr>
<td>pCVS32</td>
<td>5.3-kb HpaI fragment of pSVS11 cloned into pUC28; contains genes <em>xylC</em> (3’ end) and <em>MABN</em> (5’ end)</td>
<td>This study</td>
</tr>
<tr>
<td>pCVS42</td>
<td>5.5-kb HpaI fragment of pSVS10 cloned into pUC28, contains genes <em>xylS</em> (5’ end), <em>xylR</em>, interoperonic area, <em>xylU</em>, and <em>xylW</em> (5’ end)</td>
<td>This study</td>
</tr>
</tbody>
</table>

*a* Rif⁺, Sm² and Gm², resistance to rifampin, streptomycin, and gentamycin, respectively. His⁺, histidine auxotroph. For other abbreviations, see Table 1, footnote a.
presence of C2,3O activity which was indicative for the presence of plasmids for toluene catabolism.

**Isolation of plasmid DNA.** Recombinant plasmid DNA molecules were isolated from *E. coli* DH5α by using the method of Holmes and Quigley (46). If necessary, a neutral phenol-chloroform extraction step was introduced before precipitation of the plasmid DNA. Plasmids obtained from *Pseudomonas* and used for the restriction analysis were isolated by using a scaled-up variant of the alkaline sodium dodecyl sulfate lysis protocol of Birnboim and Doly (7) with the following modifications. After precipitation with isopropanol, plasmid DNA was dissolved in a solution of 10 mM Tris-HCl-1 mM EDTA (pH 8.0) containing 10 µg of proteinase K per ml, incubated at 50°C for 1 h, and treated with a solution of 1% cetyltrimethylammonium bromide-0.7 M NaCl to remove polysaccharides (5). The resulting supernatant was extracted once with an equal volume of neutral phenol-chloroform and once with chloroform. Plasmid DNA was again precipitated with isopropanol, and the pellet was rinsed with 70% ethanol, briefly air dried, and dissolved in 10 mM Tris-HCl (pH 8.0) containing 20 µg of pancreatic DNase-free RNase per ml. Large-scale isolation of highly purified plasmid DNA was performed by using the method of Hansen and Olsen (16), followed by CsCl-ethidium bromide (EtBr) gradient density centrifugation (46). Plasmids refractory to preparative isolation were visualized by using the in-well cell lysis method of Eckhardt (10) as described by Plazinski et al. (40) with minor modifications.

**DNA manipulations.** Transformations in *E. coli*, restriction enzyme digestion, and other DNA manipulations were carried out according to established procedures (46). Restriction enzymes and other DNA-modifying enzymes were purchased from Amersham Life Science (Little Chalfont, Buckinghamshire, United Kingdom) and used according to the specifications of the manufacturer.

**Southern hybridization.** Southern hybridization was performed essentially as described by Ravatn et al. (43, 44). *E. coli* strains carrying the cloned xyl genes of pWW0 were used to generate the following probes for hybridization: a 7.4-kb EcoRI fragment containing xylXYZLTEGFJQ, an 0.8-kb BspHI fragment with an internal portion of xylE, a 1.6-kb SmaI fragment with xylK and xylJ, a 3.0-kb SmaI fragment with xylR and xylS, and a 1.5-kb BglII fragment with xylR only. Gene probes for xylUW and xylA were synthesized as described by James and Williams (24) by PCR with the following primers: xylUW-for, 5'-TTC AGA TTG GTT GCT TTC GCC; xylUW-rev, 5'-GCT CTT TTG TTT CCC GCA TAA; xylA-for, 5'-AAG CGA AGA GCG GAA C; and xylA-rev, 5'-TTT TGG CCG CAA GAC.
GAT. A 20-kb HpaI/XbaI fragment of pSVS13 was used as a probe for testing DNA homologies to the plasmid region outside the xyl catabolic genes.

Restriction fragment length polymorphism (RFLP) analysis of plasmid DNA. Plasmid DNAs isolated by CsCl-EtBr density gradient centrifugation from each selected *Pseudomonas* isolate were digested with EcoRI, and the fragments were separated through an 0.85% agarose gel in TAE buffer (0.04 M Tris-acetate, 0.01 M EDTA [pH 8.0]) at 4°C for 14 h at 30 V without EtBr. Gels were stained with EtBr after being run and were photographed. Digitally recorded banding patterns were analyzed with the program RFLPscan (Scanalytics, Billerica, Mass.). Fragment sizes were calculated by comparing their electrophoretic mobilities with those of 1- and 5-kb markers (Bio-Rad Laboratories AG, Glattbrugg, Switzerland) and were inspected manually to avoid data misrepresentation due to any image distortion. The presence (designated 1) or absence (designated 0) of bands of the same size in a digest profile was scored and used to produce a two-dimensional rectangular data matrix of binary codes for each of the plasmid digests; this matrix was subsequently used in cluster analysis. Bootstrapping with the program SEQBOOT from PHYLIP (11) was performed to generate 100 data sets. Clustering analysis was performed on the multiple data sets by using the subroutine DOLLOP from PHYLIP with randomized input order.

ARDRA and REP-PCR genomic fingerprinting. Amplification of a nearly full-length 16S rDNA was performed by PCR with conserved eubacterial primers 16S 6F (5'-GGA GAG TTA GAT CTT GGC TCA G-3') and 16S 1510 (5'-GTG CTG GAG GGT TAC TCT GTT ACG ACT-3') (49). Repetitive sequence-based PCR (REP-PCR) genomic fingerprinting was carried out as described previously (52) with primer pair REP1R-1 (5'-III ICG ICA TCI GGC-3') and REP2-1 (5'-ICG ICT TAT CIG GCC TAC-3') and primer pair ERIC1R (5'-ATG TAA GCT CCT GGG GAT TCA C-3') and ERIC2 (5'-AAG TAA GTG ACT GTG GTG GTG AGC G-3') on total genomic DNA isolated from each *Pseudomonas* strain by the method of Marmur (32). REP-PCR products and 16S rDNA PCR copies, the latter digested with *DdeI*, *HaeIII*, or *Hinfl*, were separated on a 1 or 2% agarose gel, digitally recorded, and analyzed as described above. For one representative of each amplified 16S rDNA restriction analysis (ARDRA) class, the amplified 16S rDNA fragment was cloned into the pGEM-T Easy vector (Promega Corporation) and sequenced on a single strand only.

Amplification of *xylA* and *xylUW* by PCR. The complete *xylA*-*xylUW* area was amplified by using PCR with primers *xylA*-for/*xylA*-rev and *xylUW*-for/*xylUW*-rev (see above). A new set of primers to amplify *xylA* from plasmids pSVS11 and pSVS15 was designed in this study based on the sequence of the *xylA* area of pSVS11: *xylA11*-for, 5'-CTG
AAA AGG CCC AAG GAT AAC T; and xylA11-rev, 5'-CCA GAG CTT TTG CGG GAT GAA CTA. Purified plasmid DNA was used as a template in the PCR, and the reaction conditions were those described by James and Williams (24).

Sequence analysis. DNA sequencing was performed on an automated DNA sequencer (model 4200IR2; LI-COR Inc., Lincoln, Nebr.) as described elsewhere (43). Databases were searched for homologous gene sequences by using the BLAST program (2). Multiple sequence alignments were created by using the Genetics Computer Group program PILEUP (9) and further refined manually in WORD (Microsoft Corp.). A total of 100 multiple data sets of each alignment were generated by bootstrapping as described above. The clustering of sequences was done by using the programs PARSIMONY and NEIGHBOR from PHYLIP. Phylogenetic inferences were also evaluated by using CLUSTAL V enclosed within the Lasergene package (DNASTAR Inc., Madison, Wis.).

Nucleotide sequence accession numbers. The sequences obtained in this study are available in GenBank under accession numbers AF251321 to AF251332 (xylA), AF251333 to AF251337 (16S rDNA), and AF251338 to AF251343 (xylL and xylT).
RESULTS

Isolation and identification of toluene degraders. For all soil samples, except for the park soil without a record of oil contamination, enrichments on m-toluate grew to a visible density within 24 h. Fast-growing colony types, which appeared on TMA that had been inoculated with serial dilutions of the enrichments, were selected and further purified. This process resulted in a total of 20 strains (Table 1) consisting mostly of two different strains from each location. All strains were able to grow on toluene, m-xylene, and benzoate and demonstrated inducible C2,3O activity. None of them utilized phenol, whereas two strains utilized salicylate. Growth on salicylate, however, did not induce C2,3O activity. On the basis of morphological and physiological properties (data not shown), all strains were classified as species of the genus *Pseudomonas* according to *Bergey's Manual of Systematic Bacteriology* (30). The classification was further confirmed by ARDRA and by 16S rDNA gene sequencing.

Restriction analysis of the 16S rDNA fragments amplified by PCR with *Ddel*, *Hae*III, and *Hinf*I, including as a reference 16S rDNA derived from *Pseudomonas putida* strain mt-2 (54), divided the 20 isolates into five subgroups (Fig. 1). DNA sequences determined on one strand of cloned 16S rDNA amplification products derived from one strain of each subgroup were indeed most homologous to those of other *Pseudomonas* species (Fig. 1). This result indicated that the range of bacterial hosts carrying genes for the catabolism of toluene and xylene via a meta cleavage pathway was limited to species of the genus *Pseudomonas* belonging to the proposed "*P. fluorescens* intragenic cluster" (36). Nevertheless, on the strain level, the majority of the strains were not closely related, as judged from the dissimilarity of their REP-PCR and enterobacterial repetitive intergenic consensus-PCR patterns (data not shown).

Plasmid location of genes involved in toluene catabolism. Analysis of plasmid profiles in our *Pseudomonas* isolates revealed the presence of extrachromosomal circular covalently closed DNA elements of high molecular weight in 17 strains. The plasmids were named pSVS3 through pSVS24, and their sizes were estimated to range from 78 to 200 kb (Table 1). No large plasmids were detected in strains SV1, SV2, and SV25 by any of the plasmid visualization techniques applied. From analogy to the TOL plasmids (4), it seemed reasonable to assume that the genes for toluene catabolism would be located on the pSVS plasmids.
FIG. 1. Phylogenetic tree based on a comparison of 1,499 positions in the 16S rDNA sequences of *Pseudomonas* sp. SV strains, representative of each ARDRA subgroup, and eight neighbouring sequences from GenBank database (accession numbers are shown in parentheses): *P. putida* MnB1 (U70977), *P. putida* mt-2 (L28676), *P. plecoglossicida* (AB009457), *P. mandelli* (AF058286), *P. graminis* sp. nov. (Y11150), *P. fuscovirens* JCM 2400T (AB021378), *P. meliae* MAFF 301463T (AB021382), and *P. rhodesiae* (AF064459). The 16S rDNA sequences of *Sphingomonas paucimobilis* strain UT26 (AF039168) and *E. coli* (A14565) were used as outgroup. Groups of SV strains determined by ARDRA were as follows: (i) SV1, SV2, SV4, SV20 and SV25; (ii) SV9, SV10, SV11, SV12, SV13, SV15, SV16, SV17, SV19, SV20, SV22, and SV24; (iii) SV3; (iv) SV5 and (v) SV23 (bold face names appear in the tree). Tree was constructed by using Clustal V. The scale below the tree indicates the sequence distances as number of substitutions per 100 nucleotides.

Thus, we tested this assumption in a series of plasmid curing and conjugation experiments. Repeated cultivation on benzoate produced derivatives for all strains which were unable to utilize *m*-toluate, toluene, or *m*-xylene and which lacked C2,3O activity, except for SV1, SV2, SV22, SV23, and SV25. Of those remaining five, Tol−, C2,3O-negative (C2,3O−) derivatives could be obtained only for strains SV1, SV2, and SV23 and not for strains SV22 and SV25 by using mitomycin C. Both benzoate and mitomycin C are known to be effective in curing plasmids involved in toluene degradation in *Pseudomonas* (31, 38, 39, 56, 57). Like other strains bearing TOL plasmids, the Tol−, C2,3O− derivatives obtained in our study all demonstrated higher growth rates on benzoate than did their wild-type counterparts and finally displaced these from the benzoate-grown cultures (data not shown).
Plasmid analyses of the wild-type and Tol⁻, C2,3O⁻ derivative strains (Table 3) clearly indicated large-scale rearrangements in or complete disappearance of the plasmid DNA in the derivatives. This result suggested that the pSVS plasmids were involved in toluene degradation. Interestingly, the 120-kb plasmid of strain SV22 (pSVS22) was stable, and no Tol⁻ derivatives could be produced. However, since pSVS22 demonstrated structural similarity to the other plasmids (see below), we believe that this plasmid encoded toluene catabolism. Since no large plasmid replicons could be detected in strains SV1, SV2, and SV25, since cultivation on benzoate did not lead to the appearance of C2,3O⁻ derivatives, and since the Tol⁺ phenotype could not be transferred by conjugation, we concluded that in these three strains the genes responsible for toluene degradation are located on the chromosome.

To test the presence of additional markers on the pSVS plasmids, such as resistance to antibiotics or Hg²⁺, growth of the wild-type strains and that of their plasmid-free derivatives or transconjugants on agar plates supplemented with the appropriate antibiotics or with HgCl₂ were compared (35). Although many strains were highly resistant to ampicillin, tetracycline, kanamycin, chloramphenicol, streptomycin, or nalidixic acid (Table 1), none of these markers could be associated with the presence of pSVS plasmids. In contrast, two of five strains resistant to Hg²⁺, i.e., SV16 and SV20, seemed to carry mercury resistance genes on their plasmids, since the mercury resistance marker could be cotransferred together with the genes involved in toluene degradation to a new host (Table 3). From filter mating experiments with the different Pseudomonas SV strains and a Rif⁺ Sm⁺ derivative of P. putida strain AC13 as a recipient, 12 m-toluate-degrading transconjugants were obtained. The transfer frequencies varied greatly: from 1.0 x 10⁻⁸ to 2.0 x 10⁻³ transconjugants per donor cell (Table 3). All the transconjugants possessed the same catabolic phenotype and carried a plasmid of the same size as the respective donor strain, as judged by the electrophoretic mobility of the intact plasmid molecules (data not shown). This result suggested that during plasmid transfer, no substantial plasmid rearrangements took place. No transconjugants were obtained when Pseudomonas sp. strains SV1, SV2, and SV25 were used as donors.

Incompatibility of pSVS plasmids. All the pSVS plasmids, except for pSVS15, were stably comaintained with pMG18 and RP1, suggesting that they belong to plasmid incompatibility groups different from IncP1 and IncP9. Plasmid pSVS15 was stably comaintained with RP1 but not with pMG18, suggesting that the plasmid replicon of pSVS15 is of the IncP9 type. This notion was further confirmed by sequencing of a PCR-amplified IncP9-specific region from pSVS15 (A. Greated, personal communication). Interestingly, in a
TABLE 3. Characteristics of loss of catabolic function after repeated cultivation of toluene-degrading strains on benzoate and frequencies of conjugational transfer

<table>
<thead>
<tr>
<th>Pseudomonas sp. strain</th>
<th>Characteristics of Ben(^+), Tol(^-) derivatives</th>
<th>Frequency of conjugational transfer(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV1 and SV2 (no large plasmids detected)</td>
<td>Obtained after mitomycin C treatment</td>
<td>&lt;1.0x10(^{-8})</td>
</tr>
<tr>
<td>SV3 (pSVS3)</td>
<td>Loss of the entire plasmid</td>
<td>5.3x10(^4)</td>
</tr>
<tr>
<td>SV4 (pSVS4)</td>
<td>Loss of the entire plasmid</td>
<td>4.2x10(^{-7})</td>
</tr>
<tr>
<td>SV5 (pSVS5)</td>
<td>Loss of the entire plasmid or 40-kb deletion</td>
<td>1.1x10(^{-6})</td>
</tr>
<tr>
<td>SV6 (pSVS6)</td>
<td>Loss of the entire plasmid</td>
<td>&lt;1.0x10(^{-7})</td>
</tr>
<tr>
<td>SV9 (pSVS9)</td>
<td>Loss of the entire plasmid</td>
<td>1.0x10(^{-4})</td>
</tr>
<tr>
<td>SV10 (pSVS10)</td>
<td>Loss of the entire plasmid, 40-kb deletion</td>
<td>~ 1.0x10(^{-7})</td>
</tr>
<tr>
<td>SV11 (pSVS11)</td>
<td>Loss of the entire plasmid or 5-, 20-, or 40-kb deletion</td>
<td>1.5x10(^{-3})</td>
</tr>
<tr>
<td>SV12 (pSVS12)</td>
<td>Loss of the entire plasmid or 35- or 45-kb deletion</td>
<td>3.3x10(^{-4})</td>
</tr>
<tr>
<td>SV13 (pSVS13)</td>
<td>Loss of the entire plasmid or 35- or 60-kb deletion</td>
<td>1.0x10(^{-4})</td>
</tr>
<tr>
<td>SV15 (pSVS15)</td>
<td>Loss of the entire plasmid</td>
<td>2.0x10(^{-3})</td>
</tr>
<tr>
<td>SV16 (pSVS16)</td>
<td>Loss of the entire plasmid or 45-kb deletion</td>
<td>&lt;1.8x10(^{-8}) (^b,c)</td>
</tr>
<tr>
<td>SV17 (pSVS17)</td>
<td>Loss of the entire plasmid</td>
<td>1.0x10(^{-8})</td>
</tr>
<tr>
<td>SV19 (pSVS19)</td>
<td>Loss of the entire plasmid or 35-, 60- or 70-kb deletion</td>
<td>3.5x10(^{-7})</td>
</tr>
<tr>
<td>SV20 (pSVS20)</td>
<td>20-kb deletion</td>
<td>2.9x10(^{-7}) (^c)</td>
</tr>
<tr>
<td>SV22 (pSVS22)</td>
<td>No mutants obtained</td>
<td>&lt;5.5x10(^{-9})</td>
</tr>
<tr>
<td>SV23 (pSVS23)</td>
<td>Loss of the entire plasmid after treatment with mitomycin C</td>
<td>&lt;1.0x10(^{-8})</td>
</tr>
<tr>
<td>SV24 (pSVS24)</td>
<td>40-kb deletion</td>
<td>&lt;1.0x10(^{-8})</td>
</tr>
<tr>
<td>SV25 (no large plasmid detected)</td>
<td>No mutants obtained</td>
<td>&lt;1.0x10(^{-9})</td>
</tr>
</tbody>
</table>

\(^a\) The frequency of transfer of pSVS plasmids was calculated as the ratio of the number of transconjugants detected after mating to the number of donor cells (P. putida strain AC13) added to the crossing mixture.

\(^b\) Plasmid was not able to transfer by itself, but could be mobilized into new host by the plasmid RP1.

\(^c\) Mercury resistance genes were co-transferred with the pSVS plasmid.
few instances, colonies of *Pseudomonas* sp. strain SV15 to which pMG18 was transferred retained the catabolic phenotype of the parent strain SV15. Analysis of the plasmid content of these colonies indicated restriction fragments of only pMG18, suggesting complete elimination of pSVS15 but an apparent integration of the toluene catabolism genes from pSVS15 into the chromosome.

**Organization of toluene catabolism genes on plasmids and their homology to *xyl* genes of pWW0.** We purified plasmid DNAs from 12 strains and analyzed the plasmids by Southern hybridization for the presence of DNA homologous to the *xyl* genes of pWW0. Hybridization of the *Eco*RI-digested DNA of the pSVS plasmids with a 7.4-kb DNA probe containing the *xylX* through *xylQ* genes of pWW0 (20) showed clear signals, indicating that indeed *xyl*-homologous genes were present on the pSVS plasmids. However, we discovered three types of hybridization patterns, one exemplified by pSVS11 and pSVS15, a second by most of the other isolated plasmids, and the third unique for pSVS16 (Fig. 2).

Subsequent hybridization with the other *xyl* gene probes to plasmid DNAs digested with the restriction enzymes *Eco*RI, *Hpa*I, and *Xho*I further refined the presence and localization of *xyl*-related genes on the pSVS plasmids. From this analysis it became apparent that several pSVS plasmids carried two *meta* cleavage gene clusters (Fig. 3). The positions of restriction sites within the *xyl*-homologous regions were verified by cloning into pUC18 the 1.9-, 5.5-, 9.0-, and 10.2-kb *Eco*RI fragments from both pSVS11 and pSVS13 and the 2.3-, 3.3-, and 5.1-kb *Eco*RI fragments from pSVS13, covering genes of both *meta* pathway operons (Fig. 3). Restriction analysis showed that plasmids pSVS11 and pSVS15 (designated type C) carried a similar *meta* cleavage gene cluster (designated *meta* operon I), followed by the *xylS* and *xylR* genes and by the upper pathway operon (designated upper operon II).

The other 10 plasmids, referred to as type A and type B, carried *xyl* gene clusters similar to those of type C but, in addition, another *meta* pathway operon (designated meta operon II) and two extra copies of *xylS*-homologous sequences. One of these, *xylS*2, was situated at the end of *meta* operon II, but a second, *xylS*3, was not directly associated with the other *xyl* genes (Fig. 3). The organization of *xyl* genes found in plasmids of type A and type B was very similar to that found in TOL plasmid pWW53, isolated in North Wales, United Kingdom (26). Plasmid pWW53 has been extensively studied; therefore, the organization of the *xyl* genes and the physical map of the complete plasmid (3, 15, 27, 38) could be compared in detail to those of the pSVS plasmids.
FIG. 2. (A) Selected pSVS plasmid DNAs digested with EcoRI. (B) Autoradiogram of hybridization with 7.4-kb EcoRI fragment of pWW0 containing the meta pathway operon genes xylXYZLTEGFJQ. Molecular size markers are indicated on the left in kilobase pairs (kb). Arrows on the right indicate sizes of fragments hybridizing to the probe. Note the absence of the 2.3-kb, 5.1-kb and 3.3-kb fragments in the digests of pSVS11 and pSVS15, which are characteristic for the presence of a second copy of the meta pathway operon. The 9-kb hybridizing fragments in pSVS11 and pSVS15 were also slightly larger than those in the other pSVS plasmids (marked by the white arrows on the autoradiogram). Note also the absence of the 9.0-kb, 5.5-kb and 1.9-kb fragments from pSVS16 and appearance of a new fragment with a size of 17 kb. Digital image recorded on Gel Print 2000I (MWG-Biotech), stored as TIFF file and displayed in Canvas 3.5.5 (Deneba Software, Miami, Florida, USA).
Conservation of xyl gene order among pSVS plasmids of types A and B. Complete physical maps for plasmids of types A and B (pSVS4, pSVS9, pSVS10, pSVS12, pSVS13, pSVS16, pSVS19, pSVS20, pSVS22, pSVS23, and pSVS24) and pWW53 were generated by digestion with SpeI and XbaI and orientation of those sites with respect to the restriction maps for EcoRI and HpaI. The relative positions of the various xyl-related genes on the pSVS plasmids were confirmed by hybridization with corresponding xyl gene probes. From the sizes and the presence of all HpaI/SpeI restriction fragments smaller than 17 kb, which cover the whole area of the catabolic genes and the interoperonic region (Fig. 4 and 5), we deduced that all plasmids, except for pSVS16, pWW53, and pSVS23, were the same in this area. Displacement of one 17-kb fragment flanking the area of the catabolic genes from pSVS23 (Fig. 5) could be explained by assuming the loss of the single HpaI site from meta operon II and a shifted position of the rightmost HpaI site. Displacement of the other 17-kb fragment could be explained by assuming the absence of the leftmost SpeI site, which is also absent from pSVS13 and pSVS22 (Fig. 4). Plasmid pSVS16 had the same restriction sites as pWW53 in the catabolic gene area, except for one XbaI site (Fig. 4). A total of 22 restriction sites were conserved among the catabolic operons of type B pSVS plasmids (Fig. 3 and 4), which were slightly different in pSVS16 and pWW53. The main other variability among the plasmids of types A and B occurred outside the catabolic gene area (Fig. 4).

Specific features of pSVS11 and pSVS15. In contrast to the type A and B plasmids, pSVS11 and pSVS15 carried single upper and meta pathway gene clusters (Fig. 3). These resembled the operons of the type A and B plasmids, although the upper pathway operon was located about 1.0 kb farther from xylR than in the type A and B plasmids (Fig. 3). The positions of restriction sites within the upper pathway operons of both pSVS11 and pSVS15 were also slightly different from those in the type A and B plasmids (Fig. 3). However, the meta pathway operons of pSVS11 and pSVS15 were indistinguishable in the physical map from meta operon I of the type B plasmids (Fig. 3). It is noteworthy that the area of homology in pSVS11 and pSVS15 seemed to be extended at least 5 kb beyond the beginning of the meta pathway operon, as judged from the presence of a similar XhoI fragment of 7.2 kb (Fig. 3). Outside the catabolic gene region, however, pSVS11 and pSVS15 did not share similar restriction sites and seemed to differ extensively from one another and from the type A and B plasmids (Fig. 2 and 6).
FIG. 3. Comparative organization of xyl genes on the different pSVS plasmids and pWW53. Locations of individual genes are marked with hatched and cross-hatched arrows, the directions of which indicate the direction of transcription. The abbreviations for restriction enzyme sites are as follows: E, EcoRI; H, HpaI; X, XhoI. Only XhoI recognition sites within and close to meta pathway operons are shown. Restriction sites conserved among meta pathway operons are underlined and in bold face, those absent from pSVS16 are marked by asterisks, and the one in parentheses is present only in pSVS15. The area between upper operon I and meta operon II and xylS3 is indicated by horizontal broken lines to indicate longer distances (see Fig. 4 for exact distances). Vertical broken lines connect identical restriction sites within and close to the upper pathway operons. Note a 1.0-kb longer spacing between upper pathway operon I and meta pathway operons I in pSVS11 and pSVS15, than between those in plasmids of types A and B. Doubly shaded bars below the restriction maps indicate the EcoRI fragments cloned from pSVS13; those cloned from pSVS11 are shown by hatched bars. The 2.1-kb XhoI fragments which were cloned from pSVS11, pSVS15, pSVS13, and pWW53 for sequencing are indicated by a thin black bars. Thick black bars indicate the HpaI fragments cloned into pCVS31, pCVS32 and pCVS42 (Table 2). A 370-nt region upstream of xylC in plasmids of types A and B is 99% identical to the ntnW gene of P. putida TW3 (93% and 96% identical to xylW in pWW0 and pDK1, respectively), but is preceded by a region without any homology to xylW (open box). The open arrow for the xylA-sequence of pSVS11 and pSVS15 depicts its distinct low homology to the other xylA sequences, as opposed to the surrounding xylC, xylM, xylB, and xylN sequences.
FIG. 4. Aligned restriction maps of selected plasmids of type A and B. Positions of individual \textit{xyl} genes are depicted as in Fig. 3. Abbreviations: H, \textit{Hpa}I; S, \textit{Spe}I; X, \textit{Xba}I. Asterisks depict \textit{Hpa}I restriction sites absent from the catabolic gene area of pSVS23. Underlined are \textit{Hpa}I and \textit{Spe}I restriction sites absent from the upper pathway operons in pSVS16 and pWW53. Filled circles mark \textit{Spe}I sites absent from pSVS13, pSVS22, and pSVS23. The black bar indicates the region of pSVS13 used for probing plasmid backbone similarities. Doubly shaded bars indicate DNA areas in the other plasmids hybridizing to the backbone probe. A putative 16-kb DNA insertion in pSVS16 compared to pWW53 is shown as a dashed bar. Note the size and the restriction site variability in the regions outside the catabolic genes.

\textbf{Relationships among type A, B, and C plasmids.} According to the RFLP patterns of the complete plasmids generated by \textit{Eco}RI digestion (Fig. 2), the pSVS plasmids and pWW53 were classified into three groups (Fig. 6). This clustering was entirely consistent with the three groups based on the organization of the \textit{xyl} genes and the positions of the restriction sites within the area of the catabolic genes (Fig. 3). However, even within the largest group of plasmids (type B), two subclusters could be distinguished (Fig. 6). Similarities among the plasmids outside the \textit{xyl} gene regions were further tested by Southern hybridization of \textit{Spe}I/\textit{Xba}I-digested plasmids with the 18-kb \textit{Hpa}I/\textit{Xba}I fragment of pSVS13 (Fig. 4). We assumed that this fragment encoded plasmid maintenance and conjugation functions, since pSVS13 was conjugative and this fragment covered the smallest contiguous region among pSVS plasmids outside the catabolic gene area (Fig. 4). All the plasmids of types A and B but not of type C showed strong hybridization under high-stringency conditions with the pSVS13 fragment (Fig. 4). Still, the overall sizes of the complete plasmids were substantially different (for example, 78 kb for pSVS13 and ca. 200 kb for pSVS4 and pSVS12).
FIG. 5. Stained agarose gel (A) and graphical interpretation (B) of the HpaI/SpeI double digest of selected pSVS and pWW53 plasmid DNAs. Molecular size markers are indicated on the left in kilobase pairs (kb). Note the similarities in restriction fragments smaller than 17 kb, which are representative for the xyl operons. One 17 kb fragment is missing from pSVS23. Displacement of three fragments (depicted by 0) by a single 7.8-kb fragment (depicted by +) in pSVS16 and pWW53 is the result of a loss of the single HpaI and SpeI sites from their upper pathway operons (see also Fig. 4). The presence of three new fragments in pSVS16 (indicated by asterisks), displacing the largest HpaI/SpeI fragment in pWW53 is thought to be due to an extra 16-kb of DNA in the plasmid backbone of pSVS16 (compare with Fig. 4). The digital image was recorded on Gel Print 2000I (MWG-Biotech), stored as TIFF file and displayed in Canvas 3.5.5.

Furthermore, large DNA areas did not hybridize with either xyl or pSVS13 gene probes and had different restriction sites (Fig. 4 and 5). From these results we concluded that the type A and B plasmids have a common plasmid backbone but that large DNA rearrangements, e.g., additions and deletions, outside the xyl gene regions must have occurred. The type C plasmids showed no strong homology in their backbones with those of types A and B, since they did not hybridize to the pSVS13 probe. In addition, pSVS15 belonged to a distinct incompatibility group, as shown above.
FIG. 6. Dendrogram illustrating the relationships among the pSVS plasmids and pWW53 based on their RFLP patterns, generated by cutting with EcoRI. Clustering analysis was performed on the multiple data set by using DOLLOP. Bootstrapping values are shown in the nodes. Note that the plasmids are clustered similarly to the arrangement based on the restriction analysis of their catabolic genes (Fig.3).

Comparison of DNA sequences of the meta pathway operons of pSVS11, pSVS13, pSVS15, pWW53, and pWW0. In order to evaluate the degree of sequence similarity among the different meta cleavage gene clusters, we cloned into pUC28 and partly sequenced the conserved 2.1-kb XhoI fragments from pSVS13 (as an example for the type B plasmids), pSVS11, pSVS15, pWW53-3001, and pWW53-3002 (Table 2). All the fragments invariably contained DNA homologous to the xyl genes of pWW0 and in the following gene order: xylL (3' end), xylT, xylE, and xylG (5' end) (Fig. 3). The 730-nucleotide 5'-end sequences were aligned together with the homologous sequence of pWW0 (GenBank accession no. M64747; positions 3844 to 4604), encoding the last 127 amino acid residues of the C-terminal end of XylL (257 amino acids; accession no. S23485) and 64 amino acids of the N-terminal end of XylT separated by a noncoding intergenic region. Maximum-parsimony analysis statistically divided the sequences into three groups, with four sequences of meta operon I clustering together (Fig. 7A) but separated from the sequences of meta operon II and the pWW0 sequence. Estimated phylogenetic distances among the four type I sequences were negligibly short (0.0 to 0.1% sequence divergence), indicating that meta cleavage gene clusters of type I,
FIG. 7. Phylogenetic comparison of selected DNAs from the different meta pathway operons and upper pathway operons of several pSVS plasmids of type A, B and C, and corresponding sequences of other TOL plasmids. (A) Alignment based on 730-nt within xylL and xylT (Fig. 3). (B) Alignment based on 395-nt at the 3’ end of xylA. From the pWW0 sequence of the xylLT region 36 nucleotides (positions 4229-4265) were left out, apparently introduced by mistake (P. Golyshin, personal communication). Trees were constructed by using Clustal V. The scales below the trees indicates sequence distances as the number of substitutions per 100 nucleotides. Only bootstrapping values above 50% occurring in 100 replicate trees, obtained by using the PARSIMONY program, are shown in the nodes.

although present on different plasmids, are virtually identical. However, this group of sequences was quite distant from the pWW0 sequence (4.6 to 4.7% divergence) and even more distant from the meta operon II sequences (14.8 to 15.4% divergence). The level of divergence within sequences of the meta operon II was about 0.4%.

Comparison of xylA sequences. By using PCR with purified plasmid DNAs, we amplified xylA gene fragments of 11 pSVS plasmids (omitting pSVS24) and pWW53. Sequences of 395 nucleotides from the 3’ end of each xylA open reading frame were compared with the corresponding DNA sequences of pWW0 (accession no. D63341), pDK1 (accession no. AF019635), and ntnA (accession no. AF043544). Phylogenetic analysis divided all sequences into three groups (Fig. 7B), similar to those based on the restriction analysis. The xylA sequences of all pSVS plasmids of type B appeared to be 99.5 to 100% identical to each other, to xylA of pWW0, and to ntnA from the chromosome of P. putida strain TW3 (24). In contrast to the differences between the meta pathway operon sequences of pWW0 and type A and B plasmids, xylA of pWW0 was virtually indistinguishable from that of the type B plasmids (Fig. 7B). The xylA sequences of pWW53, pSVS16, and pDK1 formed another cluster, characterized by 3.0 to 3.5% sequence divergence from the first xylA group and by 0.0
to 0.3% divergence within the group. Plasmids of type C, pSVS11 and pSVS15, carried *xylA* sequences sharing only 71.9 to 73.7% nucleotide identity to the other *xylA* sequences, although at the predicted amino acid level the degree of similarity was higher. For example, XylA of pWW0 shared 84% and 83% amino acid residues with XylA of pSVS11 and XylA of pSVS15, respectively. The copies of *xylA* found in pSVS11 and pSVS15 were 3.5% different from each other (Fig. 7B).

Interestingly, the lower degree of homology of the upper pathway operons of type C plasmids was restricted only to the *xylA* gene. Nucleotide sequence homologies upstream of *xylA*, as determined from pSVS10 (type B) and pSVS11 (type C), approached 99% identity with each other and with *ntnC* (accession AF043544) and about 95% identity with either pWW0 (accession no. D63341), pDK1 (accession no. AF019635), or *tmbC* (accession no. U41301). Downstream of *xylA*, the sequences of pSVS10 and pSVS11 were about 95% identical to one another and shared 99 and 96% nucleotide identities with *ntnB*, respectively. These results suggested that the *xylA* sequences of the type C plasmids had an origin different from that of the surrounding upper pathway operon sequences.

Interestingly, the upper pathway operon of pWW0 itself seemed to be a mosaic-like structure, too, when compared to the upper pathway operons of type A and B plasmids. Nucleotide sequences of the *xylWCMA* genes of pWW0 were about 95% identical to the corresponding regions of type A and B pSVS plasmids, of pWW53 and pDK1, and of the *ntnWCMA* genes. However, downstream of *xylA*, the pWW0 sequence shared only 73 to 75% nucleotide identity with the other upper pathway operon sequences. Therefore, we propose that the different existing upper pathway operons have undergone several clear DNA recombinations, much more pronounced than those in the *meta* pathway operons.
DISCUSSION

The present study describes the molecular diversity of plasmids involved in toluene and xylene catabolism via the *meta* cleavage pathway in *Pseudomonas* strains (mostly) isolated from oil-contaminated sites near Minsk, Belarus. On all the isolated plasmids, the presence of *xyl* genes with different degrees of homology to *xyl* genes of other, previously described TOL plasmids could be demonstrated (1, 4 [and references within], 8, 13, 26, 28, 29, 31, 37, 54, 55, 57, 58). Two main types of gene organization of the *xyl* meta cleavage and upper pathway operons could be distinguished on the different pSVS plasmids which, however, were similar to those of some TOL plasmids isolated from other geographical regions, i.e., pWW53 in North Wales (26) and pWW102 in The Netherlands (1).

Our data suggest that *xyl*-type genes occur exclusively in members of the genus *Pseudomonas*, as was noted by others previously (4). From all contaminated soil samples but also from samples of soil with no record of oil contamination, *Pseudomonas* strains carrying *xyl*-type genes grew fastest on plates with *m*-toluate as the sole carbon source. We did not investigate the potential bias of selecting for particular types of *xyl* gene organization introduced by plating the bacteria from the soil samples on *m*-toluate. It can be expected that the conjugative plasmids on which the *xyl* genes reside will travel to hosts other than *Pseudomonas*, as was previously suggested by others (23, 42). The reason that other host bacteria carrying TOL plasmids were not retrieved in our study might be that the expression of the *xyl* genes in those bacteria is poor. A comparison of host and plasmid specificities in our study did not indicate any preferences for host-plasmid combinations, suggesting at least an "unlimited" distribution of the pSVS plasmids among the *P. fluorescens* subgroup.

A second general notion which can be drawn from our work is that the evolution of *xyl*-type genes proceeds almost exclusively in association with plasmids. Only in one strain (SV25) and perhaps in two others (SV1 and SV2) could the genes involved in toluene and xylene catabolism be located on the chromosome. The main types of evolutionary mechanisms which might have contributed to the present-day configuration of *xyl* genes in the pSVS plasmids seem to be the following: (i) recombination of complete *xyl* operons, (ii) interoperonic recombinations, and (iii) genetic drift within *xyl* coding sequences.

On all our plasmids, we found only complete copies of the *meta* cleavage operon and the upper pathway operon. Incomplete copies or physically separated *xyl* gene units smaller than the known operons were not present among the pSVS plasmids and have not been
reported for other naturally occurring TOL plasmids, except for pWW15 (37). One exception is plasmid pNL1 of Sphingomonas strain F199, on which several xyl-homologous genes were detected in rather different configurations (45). Interestingly, most pSVS plasmids carried two meta cleavage pathway operons, which were similar but not identical to one another. This type of configuration also has been detected previously in other TOL plasmids, such as pWW53. The configuration of two meta operon copies is essentially unstable, since the presence of large homologous DNA regions in direct orientation must provoke deletions via legitimate recombination. Indeed, deletion mutants of this type were detected during growth of the SV strains on benzoate (data not shown), as for pWW0 (34), pWW53 (38), pWW15 (37), and pDK1 (31). On the other hand, recombination between the two meta operons would lead to a loss of the upper pathway operon and of xylS and xylR, a result which is clearly disadvantageous for cells growing on hydrocarbons or m-toluate. It is not clear if the presence of two meta operon copies per se will have a selective advantage for cells, such as higher expression of the meta pathway enzymes or formation of heteromultimeric enzymes.

The finding of structurally dissimilar plasmid backbones and different incompatibility groups among the pSVS plasmids but still similar xyl operons argues for independent "movement" of the xyl operons and independent acquisition of the xyl genes by different plasmid vehicles. In several TOL plasmids, such as pWW0, pDK1, pWW53, and pWW15, the xyl operons are enclosed within a large transposable element (27, 37, 47, 50, 51), which might be the main mechanistic element for distributing the xyl genes to other plasmid replicons. However, transposable elements have not yet been described for the pSVS plasmids. The fact that the xyl operons on type C plasmids (pSVS11 and pSVS15) and those on type A and B plasmids were so alike argues for their relatively recent acquisition and/or distribution, even though smaller insertions (such as within the xylUW region in type B plasmids) seem to have taken place since then. Upper and meta cleavage pathway operons are not necessarily distributed or acquired together. This notion became apparent from the high percentage of identity between the ntn genes involved in nitrotoluene degradation in P. putida strain TW3 and the xyl upper pathway operon genes but an apparent lack of the meta pathway operon genes in strain TW3 (24).

On a smaller scale, we found evidence that insertions, deletions, and recombinations occur within xyl operons. This conclusion was obvious from a comparison of upper pathway operons. The xylA gene in plasmids pSVS11 and pSVS15 had clearly lower sequence homology to other known upper pathway operons with respect to its surrounding sequences. This finding indicates insertion of or recombination with an upper pathway operon containing
a different \textit{xyl}A gene. Strangely enough, the borders of this aberrant DNA sequence enclose the \textit{xyl}A open reading frame quite exactly. The origin of this \textit{xyl}A sequence is undefined, since it had only moderate sequence identity to \textit{xyl}A sequences of pWW0, pDK1, and pWW53, and to \textit{ntn}A. We also concluded that the upper pathway operon in pWW0 represents a hybrid structure in comparison to those in pWW53, pDK1, and the pSVS plasmids.

Finally, a small insertion seemed to have occurred in the \textit{xyl}W open reading frame in type A and B plasmids (Fig. 3). Despite our hypothesis that the \textit{xyl} operons were distributed among the pSVS plasmids relatively recently, the DNA sequences indicated some genetic drift (Fig. 7). Most coherent were plasmids of type B, which was clearly distinguishable from the other types and, therefore, may be endemic for the Minsk area. In contrast, pSVS16 was more similar to pWW53 and thus may have originated from an area in which pWW53 is endemic and been transported to Belarus.

Taken together, the molecular diversity of the \textit{xyl} genes on the pSVS and other TOL plasmids indicates a dynamic evolution even for such an established pathway as toluene and xylene degradation, involving various mechanisms. Because of such different mechanisms, it becomes practically impossible to trace the evolutionary history of individual \textit{xyl} operons. Since substantial variability in the plasmid backbones of pSVS plasmids was observed as well, we can conclude that plasmid vectors act as recipients for various kinds of "junk" and useful DNA. Due to permanent oil-related human activities, selective environments are created in which bacteria with various different combinations of \textit{xyl} operons can proliferate.

\section*{ACKNOWLEDGMENTS}

We thank G. Jacoby (Lahey Hitchcock Clinic, Burlington, Mass.) for plasmid pMG18 and A. Boronin (Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow Oblast, Russia) for \textit{P. putida} strain AC13.
REFERENCES


CHAPTER 3

Characterization of two alternative promoters for integrase expression in the clc genomic island of Pseudomonas sp. strain B13

Accepted for publication in Mol. Microbiol. by Vladimir Sentchilo, Alexander Zehnder and Jan Roelof van der Meer
SUMMARY

Expression of the integrase of the clc genomic island in Pseudomonas sp. strain B13 was studied by using single cell green fluorescent protein (GFP) reporter technology. The clc genomic island is an integrative and conjugative element (ICE) of 105 kb in size, which encodes metabolism of 3-chlorobenzoic acid. Two copies of the clc element are integrated in genes for glycine tRNA, but in a small proportion of individual cells (5-15%) the element has excised to form a circular intermediate in which both ends are connected. The integrase gene (intB13) is located 202 bp from this junction site facing inwards. Fragments upstream of intB13 in the circular form and in the integrated form were fused to a promoterless gfp gene and introduced in monocopy onto the chromosome of strain B13. Quantitative GFP fluorescence measurements in individual cells of the different B13-derivatives revealed that the circular form fragment contained a strong constitutive promoter (P\text{circ}) driving intB13 expression in all cells. By using primer extension P\text{circ} could be mapped near the left end of the clc element and P\text{circ} can therefore only control intB13 expression when left and right ends are connected as in the circular form. Expression from intB13 upstream fragments from the integrated clc element was weaker than that from P\text{circ} and only occurred in maximally 15% of individual cells in a culture. A promoter (P\text{int}) could be roughly mapped in this region by using reverse-transcription PCR and by successively shortening the fragment from the 5’ end. GFP expression from 5’ deleted P\text{int} fragments also pointed to possible repressor and activator binding sites in this 232 bp area. Transposon mutants in cloned left end sequences of the clc element were selected which had lost the activation potential on the P\text{int} promoter and those which resulted in overexpression of GFP from P\text{int}. The DNA sequence of the region of the transposon insertions pointed to a relatively well conserved area among various other genomic islands. The activator mutants mapped in an open reading frame (ORF) is encoding a 175 amino acid protein without any significant similarity to functionally characterized proteins in the databases. Transposon insertions leading to GFP overexpression mapped in a 1650 bp ORF with a weak homology to parB. It is unlikely however that this protein encodes a repressor for intB13 expression, since other transposon insertions in the same ORF had no effect on GFP production from P\text{int}. Furthermore, the overexpression mutants retained their effect even when the presumed repressor binding site at the P\text{int} operator was deleted.
INTRODUCTION

The clc element is a genomic island found in *Pseudomonas* sp. strain B13 (45, 47). It has a size of approximately 105 kb and encodes among other functions the degradation of 3-chlorobenzoic acid (CBA) and chlorocatechol. The clc element is self-transmissible among Gram-negative bacteria (40, 45, 47, 56, 57, 66), making it an interesting model to study the biology of an expanding class of mobile DNA structures referred to as ‘ecological fitness genomic islands’ (20). Genomic islands and conjugal transposons have recently been classified as members of a similar group of integrative and conjugal elements (ICE) (9). ICEs integrate site-specifically into one or more target sites (genomic islands usually one) in the chromosome (or on large plasmids), can excise from that location, form a circular intermediate in which both ends are connected and can mediate conjugal transfer to a new recipient cell in which reintegration can take place (9, 20). Genomic islands often contribute to pathogenicity of the host organism, but other factors like aromatic degradation or plant symbiosis have been found as well (for comprehensive review see (21)).

Comparatively little is known about the cellular factors and possible signals regulating excision and transfer of genomic islands. Like many genomic islands the clc element is integrated into a tRNA gene, in this case the tRNA$^{Gly}$. Recombination takes place between an 18-bp sequence of the attP-site of the clc element identical to the most 3’ 18-bp from the tRNA$^{Gly}$ and this results in a displacement of this 18-bp sequence upon integration to the left end of the clc element (attL, Fig. 1). The integration site marks the right end (attR) of the element (46) which also contains the intB13 integrase gene at a distance of 202 bp from the junction site facing inwards. The overall structure of the clc element (i.e., tRNA-integrase-repeat) is common to that of a large variety of other genomic islands, such as those in the *Pseudomonas aeruginosa* isolates C and S17GM (34) or in *Yersinia* (4, 8, 43, 44), and that of the satellite bacteriophages P4 (42) and φR73 (26), and the defective prophage CP4-57 (31). Conjugal transposons tend not to have tRNA genes as integration sites but still retain the common structure (integration site-integrase-repeat) (10, 60, 62). Despite the clearly different functional genetic load of the different genomic islands, of conjugal transposons and of bacteriophages, still some of the mechanistic steps and strategies in integration and excision may thus be similar.
FIG. 1. A) Schematic presentation of the two forms of the clc genomic island, its life cycle and reactions catalyzed by the IntB13 integrase. During integration the 18-bp 3’ end (depicted by open triangles) of the target gene for glycine tRNA (tRNA$^{\text{Gly}}$) recombines with an identical 18-bp sequence (filled triangles) originating from the right (R) end of the clc element. One of the 18-bp sequences is displaced to the other end of the integrated form. Excision results in a closed junction between left (L) and right (R) ends of the element. Promoter regions of the integrase gene ($intB13$) of the integrated ($P_{int}$) and the excised circular form ($P_{circ}$) of the clc element are depicted by thin arrows showing the direction of transcription (not to proportion). B) Integrase promoter reporter cassettes jim1 ($P_{int}$-gfp) and jim4 ($P_{circ}$-gfp). Coding sequences for green fluorescent protein (gfp) and kanamycin resistance (Km-R) genes are shown as arrows (not proportional). Solid vertical bars correspond to the I- and O-ends of the Tn5 delivery system.

For the clc element it is the IntB13 integrase which is responsible for integration of the element into the target site (46). Functional studies on related integrases (11, 35, 38, 48, 64) give good reason to assume that excision of the clc element is also mediated by the IntB13 integrase, although this has not been experimentally proven and other auxiliary factors (e.g., excisionase) may be needed for that process. The IntB13 integrase can be classified as a P4 type integrase with moderate (35%) amino acid identity to those (59) and is about one third larger in size than most other P4 type integrases (46). Since the integrase plays such an essential role in the cycle of excision and integration of bacteriophages and ICEs, regulation of integrase expression is of particular interest for understanding activity of those elements in relation to their host.

Most information on regulation of integrase expression comes from bacteriophages and conjugative transposons. Despite the different mode of cell to cell transfer (i.e., transduction vs. conjugation) bacteriophages have mechanistically similar steps in their life
cycle as ICEs, for example, formation of a closed circular molecule and target-site integration/excision. Therefore, regulation of integrase expression may proceed along similar strategies. In phage P4 integrase transcription can occur from two alternative promoters. One of these is a strong $\text{\textsuperscript{70}-type}$ promoter and is active immediately after infection of the cell with the phage. High integrase expression at this point is probably necessary to achieve integration of the phage into the host DNA. Upon lysogeny, transcription from this promoter is repressed (42) and integrase expression is now subject to a weak alternative promoter. It is not exactly known which factors induce this alternative promoter, except that coinfection of the cell with the helper phage P2 will lead to prophage excision ((19, 42) and references within). In the related P4-type phage CP4-57 of Escherichia coli K12 transcription activation of the integrase was described to be dependent on the $\text{alpA}$ gene product (31, 58). Two alternative promoters also control integrase transcription in the Y. pestis HPI genome island (43). In the excised and circular form of the HPI island transcription starts from a strong constitutive promoter located in the left end of the island, but which continues through the junction site into the integrase gene. Different to the situation in P4, the left end promoter of the HPI island is physically separated from the integrase gene upon integration, and integrase expression is then thought to be solely controlled by readthrough from the target tRNA gene (43). The theme of readthrough transcription in the circular excised form also takes place in the conjugative transposon Tn916, although the integrase gene in this case faces outward to the junction site. Transcriptional control of the integrase in the integrated Tn916 is thought to occur from a weak $\text{\textsuperscript{70}-type}$ promoter mapped upstream of the integrase gene (38). Upon excision and in the presence of tetracycline, however, a very long transcript (17 kb) starting at the promoter of the $\text{tetM}$ gene reads through the integrase gene across the junction site and into the transfer genes (10). Finally, expression of the integrase genes in the conjugative transposon CTnDOT in Bacteroides and the mobilizable unit NBU1, which both again face inward to their elements, was shown to be constitutive (11, 54), although excision and transfer of those elements was further dependent on a cascade of different transcription factors (60, 62).

In this work we characterized two alternative promoters for expression of the $\text{intB13}$ integrase by studying $\text{gfp}$ expression in Pseudomonas sp. strain B13 from single-copy chromosomal transcriptional fusions. Interestingly, by quantifying GFP expression in single cells rather than from the population as a whole, we discovered that expression of the integrase in the integrated form of the $\text{clc}$ element was subject to regulatory control, but
occurred only in a small proportion of cells. This population-dependent effect on integrase
expression may have not been seen in other studies which have used reporter genes for the
whole population (like gus, lacZ or luxAB). Primer extension analysis, promoter deletion
studies and reverse transcriptase polymerase chain reaction were used to further locate regions
in both promoters at which trans-acting factors could exert transcriptional control. Further
experimental evidence for transcription factors acting on integrase expression is presented by
mutational analysis of a left-end extremity of the clc element.
RESULTS

An outward facing promoter, $P_{circ}$, at the left end of the $clc$ element. The $clc$ element can occur in two different configurations: i) integrated in the tRNA$^{Gly}$ gene and ii) as a closed circular form in which both ends are connected (Fig. 1). In order to study whether both configurations would lead to different expression of the $intB13$ gene, DNA fragments upstream of $intB13$ in the integrated and in the closed circular form were tested for promoter activity. A DNA fragment of 480 bp encompassing the $attP$ region directly upstream of $intB13$ in the circular form of the $clc$ genomic island was fused to the gfp gene (jim4 construction, Fig. 1). *E. coli* strains DH5α and CC118$\lambda$pir carrying pJAMA23-jim4 and pCK218-jim4, respectively, produced green colonies on nutrient agar and appeared brightly fluorescent in UV microscopy, suggesting that this $attP$ fragment upstream of $intB13$ carried a constitutive promoter. Essentially the same result was obtained when GFP expression from the jim4 construction was studied in *Pseudomonas* sp. strain B13. Individual cells of *Pseudomonas* sp. strain B13-jim4 were brightly fluorescent both during exponential and stationary growth phase, on nutrient broth and on minimal medium with fructose or CBA as C-source (see Chapter 4, Fig. 2). Average gray values (AGV) of individual cells, a quantitative measure for GFP fluorescence, were normally distributed around a mean value of 200 AGV units during exponential growth and 350 AGV units during stationary phase. Four clones of *Pseudomonas* sp. strain B13-jim4 picked at random from the transconjugant plates all behaved alike (data not shown), implying that the nature of gfp expression from this DNA fragment was not influenced by the position of insertion in the B13 chromosome. These observations showed that expression of the integrase gene in the circular form of the $clc$ genomic island is regulated by a strong constitutive promoter (which we called $P_{circ}$), present on the 480-bp $attP$ fragment. Secondly, transcription from $P_{circ}$ did not seem to require B13- or $clc$-island-specific regulatory factors since it took place in two *E. coli* strains devoid of the $clc$ island. Thirdly, expression from $P_{circ}$ did not seem to require any particular growth phase or growth substrate.

To map the transcription initiation site within $P_{circ}$ total RNA from *E. coli* strain DH5α (pJAMA23-jim4) was isolated and reverse transcribed using primer SV004-BamHI (Table 1). A single cDNA fragment was found, from which the transcription initiation site was mapped within the ‘left end’ of the $clc$ element, i.e. 61 bp upstream of the junction between left and right ends (the junction is laying 202 bp upstream of the $intB13$ start codon) (Fig. 2).
TABLE 1. Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Corresponding promoter reporter, primer location</th>
</tr>
</thead>
<tbody>
<tr>
<td>jim1-SphI-Fw</td>
<td>5’ GTAGAGCA\textsubscript{ATGC}CCTTGCCAAGG\textsuperscript{3}^a</td>
<td>jim1, 253 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim2-PstI-Fw</td>
<td>5’ GATCT\textsubscript{GC}CAGCTTCTACTGACG\textsuperscript{3}’</td>
<td>jim2, 192 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim3-PstI-Fw</td>
<td>5’ GAGGCGG\textsubscript{CTGC}GGATGTTGC\textsuperscript{3}’</td>
<td>jim3, 40 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim4.1-Fw</td>
<td>5’ GAATCAACG\textsubscript{CTTG}CCTG\textsuperscript{3}’</td>
<td>Left end, 28 bp apart from the junction between left and right end</td>
</tr>
<tr>
<td>jim2.6</td>
<td>5’ TGGTTG\textsubscript{AAAG}CGGGCTTCC</td>
<td>147 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim6-SphI-Fw</td>
<td>5’ CCTTTTT\textsubscript{TGCATGC}ACGC\textsuperscript{3}’</td>
<td>jim6, 123 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim7-PstI-Fw</td>
<td>5’ CGGAAATCC\textsubscript{ACCTG}GAGCCAGC\textsuperscript{3}’</td>
<td>jim7, 99 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>jim9-PstI-Fw</td>
<td>5’ CAATCTAG\textsubscript{GACTG}CATG\textsuperscript{3}’</td>
<td>jim9, 65 bp upstream of the beginning of intB13</td>
</tr>
<tr>
<td>SV004-BamHI-R</td>
<td>5’ GAGG\textsubscript{ATCC}TCTAAG\textsubscript{TAATGAC}3’</td>
<td>General reverse primer for promoter reporters, 39 bp downstream of the beginning of intB13</td>
</tr>
<tr>
<td>SV004-R</td>
<td>5’ GAGGCT\textsubscript{CCTA}GTAATGAC\textsuperscript{3}’</td>
<td>Same as above, but with no nucleotides changed</td>
</tr>
<tr>
<td>Int271-R</td>
<td>5’ GACTCG\textsubscript{CCCTTG}GTAGCTG\textsuperscript{3}’</td>
<td>10 bp downstream of the beginning of intB13; used for primer extension</td>
</tr>
<tr>
<td>Ter-Fw</td>
<td>5’ CAGGAAT\textsubscript{TTTCGAGGCATGC}3’</td>
<td>Specific for gfp fusion cassette, binds immediately upstream of the promoter-gfp fusion</td>
</tr>
<tr>
<td>GFP-R</td>
<td>5’ GTATG\textsubscript{TGTGCATCACC}CTC\textsuperscript{3}’</td>
<td>Specific for gfp fusion cassette, binds 120 bp downstream of the beginning of gfp</td>
</tr>
</tbody>
</table>

\(^a\) Fw and R stand for the forward and reverse primer, respectively.
\(^b\) The BamHI, PstI and SphI restriction sites are shown in bold face font, mutated nucleotides are underlined.

The transcription initiation site on the sequence was preceded by strongly conserved \(-35, TTGCGG,\) and \(-10, TACAAAT,\) hexamer boxes separated by a 17 bp spacer which is typical for \(\sigma\)\(_{70}\) dependent promoters (22, 63).

**Integrase promoter, P\textsubscript{int}, of the integrated clc element.** In the integrated form of the clc island expression of the integrase must be directed by an alternative promoter, since the left end is displaced by the gene for tRNA\textsubscript{Gly} (Fig. 1). This promoter was assigned P\textsubscript{int}. To get an insight into the structure and functioning of P\textsubscript{int} two approaches were employed: (i) transcription initiation site mapping by reverse transcription (RT) or reverse transcription PCR (RT-PCR) and (ii) promoter deletion mapping by fusing different promoter fragments to the gfp gene and studying GFP expression in strain B13.

For primer extension total RNA was isolated from cultures of the strain B13-jim1 carrying plasmid pTCB207-6 grown on MM with 10 mM CBA until late stationary phase (64 h incubation). Strains with this plasmid were found to have the highest GFP expression from the P\textsubscript{int} promoter (see below). Unfortunately, no cDNA transcript could be detected in three
FIG. 2. Structure of the P_circ integrase promoter in the free circular form of the clc genomic island. A) Mapping of the transcription initiation site with primer extension. Lanes: RT, reverse transcription reaction with total RNA from E. coli DH5α (pJAMA23::jim4) extended with primer SV004-BamHI-R (0.5, 1.0, and 2.5 µl of the RT reaction were loaded on the gel); A, C, G and T, corresponding DNA sequence obtained with the same primer. The arrow marks position of the primer extension product. Coding strand DNA sequence of the promoter area surrounding the transcription start site is shown on the right. First transcribed nucleotide marked +1 and –10 hexamer (–10) are shown in boldface. B) Structure of P_circ at the junction between left and right ends of the clc island. Boxed are the 18 nucleotides at the right end border of the clc element homologous to 3' end of the tRNA^Gly (see also Fig. 1). Nucleotides of the transcription start (+1) and those matching -10 and –35 consensus hexamers are printed in boldface. The intB13 translation start codon is underlined and marked with the character M. Sequence of the primer SV004-BamHI-R used for RT is shown under the sequence. Note that the left end of the clc island provides all DNA elements relevant to promoter functioning. Arrows points to two remarkable inverted repeats.

independently repeated experiments neither with the IRD-800 labeled primer SV004-BamHI nor with a radioactively labeled primer int271-Rev (Table 1), despite the fact that both int271-Rev and SV004-BamHI efficiently and specifically amplified DNA in a PCR with B13 chromosomal DNA as a template. The most likely explanation for the failure to detect any specific cDNA fragment is a very low abundance of the intB13 transcript in the B13 RNA sample. For this reason, we tried to determine the transcription start site at least roughly by detecting the presence or absence of an amplicon in RT-PCR using a fixed reverse primer and differently located forward primers (jim1 through jim9, Fig. 3). RT-PCR products of the expected size were obtained with forward primers jim3, jim6, jim7 and jim9 but not with
FIG. 3. Structure of the P_{int} integrase promoter. A) Coding strand DNA sequence at the junction between the host chromosome and the right end of the clc-island is shown. Nucleotide numeration corresponds to the intB13 area sequence deposited in GenBank under Accession nr. AJ004950 (46). Grey shading indicates the host chromosome area, boxed is the sequence of the tRNA^{Glu} gene. The 18 nucleotides of the recombination site are boxed without shading. The integrase gene start codon is shown in boldface. Forward and reverse primers used for constructing P_{int}-gfp fusions are depicted by solid arrows in 5'-3' orientation above and below the DNA sequence. Restriction sites for cloning are indicated underneath the sequence with mutated nucleotides shown in bold. The two inverted repeats are indicated by arrows (IR1 and IR2). Approximate positions of the repressor- and activator-binding sites are depicted with dashed lines above the sequence. The beginning of the intB13 transcript originating from P_{int} is shown as a zig-zagged arrow. B) RT-PCR mapping of the 5' end of the P_{int}-intB13 transcript. Total RNA from stationary phase culture of B13-jim1 (pTCB207-6) grown in MM with 10 mM CBA was reverse transcribed with primer int271-R. The resulting cDNA was amplified with PCR using reverse primer int271-rev in combination with one of the forward primers (listed according to their position in the sequence) jim1, jim2, jim2.6, jim6, jim7 and jim3. Lanes: –RT, RT reaction without reverse transcriptase added; +RT, complete RT reaction; C, positive control with genomic DNA of B13-jim1 as template; M, DNA size-marker, sizes of the bands in base pairs are shown between the pannels.

jim1, jim2 and jim2.6 (Fig. 3). This confined the position of the P_{int} transcription initiation site to the area between the primers jim2.6 and jim6 or overlapping with jim6 (Fig. 3). Minor DNA bands were occasionally also observed in the reactions containing forward primers jim2
and jim2.6, as well as with primer jim4.1, which is annealing to the left end portion of the circular form of the clc island (Fig. 2). Since the amount of PCR product from these upstream primers, whenever present, was always lower than in the RT-PCR reactions containing primers annealing downstream of the proposed transcription initiation site, these minor products seem to result from intB13 RNA transcripts starting from the P<sub>int</sub> promoter. In almost all stationary phase cultures, a small percentage (5-15%, not shown) of cells carry the circular form of the clc element, which would be a constant source for P<sub>circ-intB13</sub> transcripts.

**int::gfp expression from P<sub>int</sub> 5'-deletion mutants.** In order to confirm the location of the transcription initiation site of P<sub>int</sub> and to map possible functional promoter and/or operator elements, six transcription fusions were constructed by cloning fragments upstream of intB13 in its integrated form subsequently deleted from the 5' end in front of the gfp gene (Fig. 3). Each promoter-gfp fragment was inserted in single copy into the *Pseudomonas* sp. B13 genome by mini-Tn5 transposition. Compared to GFP expression from P<sub>circ</sub> the GFP fluorescence levels of individual cells of strains B13-jim1, -jim2, -jim3, -jim6, -jim7 and -jim9 were all much lower. Cells growing on liquid MM with 10 mM CBA in batch displayed maximum GFP fluorescence levels in stationary phase (after 72 h incubation). Expression from the longest construct jim1 led to the formation of about 15% of cells with visible fluorescence. 95% percentile and top 5% average mean values were 100 and 108 AGV units, respectively (Table 2, Fig. 4). Shortening the promoter fragment, as in the construct jim2 (Fig. 3), significantly increased fluorescence values of single cells compared to jim1 (i.e., 161 vs 100, respectively) (Table 2, Fig. 4). Subsequent shortening to the size encompassed by the jim6-gfp fusion resulted in a decreased GFP expression, but remaining significantly higher than in jim1 (13% induced cells in stationary phase, 95% percentile of 120 AGV units) (Table 2, Fig. 4). Further shortening of the promoter region beyond the primer jim6 abolished gfp transcription completely as virtually no GFP was produced in B13 strains –jim7, -jim9 and –jim3 (Table 2, Fig. 4). For every promoter fragment, four independently selected clones of strain B13 were assayed, which all behaved alike with respect to GFP expression (not shown). This indicated that the position of the promoter-gfp fusion in the B13 genome played no role in the observed GFP expression, but that variations in GFP expression were due to deletions of essential promoter/operator elements. These observations, firstly, confirmed the approximate location of the transcription initiation site obtained by RT-PCR analysis, i.e., laying within the area adjacent to primer jim6 (Fig. 3), since shortening the promoter fragment to beyond jim6 abolished GFP expression. Secondly, they pointed at the possible
TABLE 2. Effect of 5’ deletions on the expression of $P_{int}$-gfp reporter gene fusions

<table>
<thead>
<tr>
<th>Reporter strain$^a$</th>
<th>Size of deletion, bp</th>
<th>95% percentile$^c$</th>
<th>GFP fluorescence Top 5% mean$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B13-jim1</td>
<td>0</td>
<td>100 (99-101)</td>
<td>108</td>
</tr>
<tr>
<td>B13-jim2</td>
<td>56</td>
<td>161 (158-164)</td>
<td>180</td>
</tr>
<tr>
<td>B13-jim6</td>
<td>114</td>
<td>120 (118-122)</td>
<td>134</td>
</tr>
<tr>
<td>B13-jim7</td>
<td>143</td>
<td>83.5 (83.3-83.8)</td>
<td>86.6</td>
</tr>
<tr>
<td>B13-jim9</td>
<td>178</td>
<td>85.2 (85.1-85.3)</td>
<td>85.9</td>
</tr>
<tr>
<td>B13-jim3</td>
<td>198</td>
<td>84.2 (84.1-84.2)</td>
<td>85.2</td>
</tr>
<tr>
<td>Detection limit</td>
<td>-</td>
<td></td>
<td>79</td>
</tr>
</tbody>
</table>

$^a$) Reporter strains *Pseudomonas* sp. B13-jim1, -jim2, -jim6, -jim7, -jim9 and -jim3 (listed according to the size of the promoter region deleted) were grown in liquid MM with 3-chlorobenzoate present at 10 mM concentration until stationary phase (72 h).

$^b$) The length of the promoter region is 232 bp counting from the $intB_{13}$ start codon (Fig. 3).

$^c$) GFP fluorescence intensity is given as the 95% percentile average gray value of all cells in the population. Values in parenthesis represent the 95% confidence intervals calculated with the R-software (see Experimental procedures).

$^d$) Values given are the arithmetic mean of the cellular average gray values for those 5% of the population with highest fluorescence intensities.

involvement of repression and activation mechanisms regulating expression from $P_{int}$, since shortening the fragment from jim1 to jim2 actually caused an increase in GFP expression (as if a repressor binding site was removed), but further shortening to jim6 again caused a decrease (as if an activator binding site was subsequently affected).

**Cloning and mapping of genes involved in regulation of the integrase expression.**

Preliminary sequence data of a 9 kb region at the left end of the clc element, containing gene homologs of phage- and plasmid-type transcription regulators (see below) led us to the idea that some regulatory factors for $intB_{13}$ expression were encoded in this area. To test this hypothesis, two adjacent *Eco*RI fragments spanning together the 9 kb left end region were separately brought into strain B13-jim1 on the broad-host-range hybrid vector pKT230-pUC28 (pTCB207 and pTCB208, Fig. 5) and mutagenized by transposon insertions. The reason for not creating knockouts directly on the clc element was that, unfortunately, strain B13 carries two copies of the clc element. Therefore, we chose to mutagenize the left end’s genes from the clc element on the plasmid pTCB207 and rely on gene dosage differences between the plasmid and the chromosomal copies to observe effects of insertional gene inactivation. The reason to use the hybrid vector was the ease of cloning in the unique *HindIII*
FIG. 4. Effect of promoter deletions on P_int transcription activity. GFP fluorescence intensities in populations of *Pseudomonas* sp. strain B13 carrying the complete P_int-gfp fusion (construct jim1) and its 5'-end deletion constructs -jim2 (∆56 bp), -jim6 (∆114 bp) and -jim7 (∆143 bp). All reporter strains were grown on MM with 10 mM CBA as a C-source for 72 h and were inspected in parallel. Shown are cumulative distribution plots of all cells in a population. Each point indicates the GFP fluorescence value of a single cell expressed as average gray value (AGV) units calculated with METAMORPH. A value of 79 AGV corresponds to the background level of the camera system (i.e., cells are dark and have no detectable fluorescence). The GFP fluorescence value at the 95% percentile of each of the population is indicated on the X-axis. The corresponding 95% confidence intervals are too narrow to be shown (Table 2). Note the increased GFP production in B13-jim2 compared to –jim1, and its attenuation in -jim6. Deletion of the 143 bp in B13-jim7 abolished GFP expression completely.

Site present on both pKT230 and pTCB177 and the necessity for a second antibiotic resistance marker (ampicillin) to avoid propagation of possible spontaneous streptomycin resistant mutants in *Pseudomonas* sp. strain B13. In addition, the origin of replication of pUC28 gave an advantage of high plasmid yield in *E. coli* required for subsequent DNA sequencing of the transposon insertion sites.

Introduction of pTCB208 carrying the most exterior 4.2 kb EcoRI fragment (Fig. 5) into strain B13-jim1 exerted no effects on the magnitude of GFP expression per cell nor on the relative fraction of induced cells in the whole population (not shown). In contrast, introduction of the plasmid pTCB207, carrying the 6.4 kb EcoRI fragment located more inwards from the left end (Fig. 5) caused measurably increased GFP production compared to B13-jim1 devoid of any plasmid (Table 3). This suggested factors to be encoded in this area which were influencing expression of the P_int promoter. Plasmid pTCB207 was then in vitro mutagenized with the tetracycline resistance insertion cassette <TET-1> in order to produce interruptions of any relevant gene functions. Twelve pTCB207 transposon mutants (Fig. 5) were again introduced via conjugation into B13-jim1 and the effects on GFP expression were
FIG. 5. Genetic organization of the 6,383-bp EcoRI fragment present in pTCB207 and location of the <TET-1> transposon insertions. A) Junction (attL) of the left end of the clc island (depicted by a solid line) with the chromosome (dashed line), with the location of the 4.2 and 6.4 kb EcoRI fragments (45). B) Genetic map of the insert in pTCB207. Relevant features of pKT230 and pUC28 adjacent to the 6.4 kb EcoRI insert are shown. Direction of transcription from the promoter of the kanamycin-resistance gene is indicated. Open arrows (drawn to proportion) mark the positions of major open reading frames (orf1-8) identified with DNASTAR. A thick outline marks the putative activator of the integrase. Location of the <TET-1> transposon insertions is depicted by triangles and numbers of the corresponding pTCB207 mutant plasmids are shown underneath. Filled triangles indicate insertions that caused “up” or “down” regulation of $P_{int}$-gfp expression. Open triangles points to insertions without measurable effect on GFP expression.

studied in late stationary phase after growing in MM with 10 mM CBA as C-source. Two clones of B13-jim1 carrying plasmids with transposon insertions in the right proximity of the 6.4 kb insert (pTCB207-41 and -46; Fig. 5) displayed significantly less GFP both in fluorescence of individual cells and in the fraction of induced cells in the population compared to B13-jim1 (pTCB207) (Table 3). Conversely, three other transposon insertions all clustering between coordinates 2 and 3 kb, as in pTCB207-6, -52 and -64 (Fig. 5), caused a significant increase of both the fraction of induced cells in the population and the levels of GFP fluorescence in individual cells compared to B13-jim1 (pTCB207) (Table 3). These data suggested that two transposon insertions (nrs. 46 and 41) had interrupted a gene for an activator of intB13 expression (resulting in lower GFP production) and three others (nrs. 6, 52 and 64) had interrupted a gene for a repressor or the one indirectly effecting intB13 expression (yielding increased GFP expression levels). The seven remaining transposon insertions in pTCB207 did not measurably change the character of $P_{int}$ expression as
TABLE 3. Regulatory effect of pTCB207 and its selected transposon insertion mutants on GFP expression from different reporter constructs in *Pseudomonas* sp. strain B13

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid</th>
<th>GFP fluorescence(^a)</th>
<th>95% percentile</th>
<th>Top 5% mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>B13-jim1*</td>
<td>-</td>
<td>97 (93-99)(^c)</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pTCB207</td>
<td>107 (105-109)</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>pTCB207-6(^c)</td>
<td>122 (119-125)</td>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTCB207-7</td>
<td>108 (107-110)</td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTCB207-41</td>
<td>97 (95-99)</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTCB207-48</td>
<td>106 (103-108)</td>
<td>119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTCB207-62</td>
<td>105 (103-107)</td>
<td>115</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B13-jim2

|              | -          | 137 (125-146) | 200 |             |
| pTCB207    | 180 (166-191)| 238        |     |             |
| pTCB207-6  | 223 (216-232)| 286        |     |             |
| pTCB207-41 | 149 (140-161)| 213        |     |             |

B13-jim6

|              | -          | 121 (119-123)| 134 |             |
| pTCB207-6  | 166 (163-169)| 198        |     |             |

\(^a\) Reporter *Pseudomonas* sp. strains B13-jim1, -jim2 and -jim6 (with or without plasmid to test) were grown in MM with 3-chlorobenzoate present at 10 mM concentration until late stationary phase (7 days).

\(^b\) As for Table 2.

\(^c\) Highlighted with bold face are pTCB207 derivatives where <TET-1> transposon insertions resulted in up- or down-regulation of P\(_{int}\)-gfp fusion.

exemplified by pTCB207-7, -48 and -62 (Table 3).

**DNA sequencing and sequence analysis.** The sequence of the DNA region present in pTCB207 was determined. This region is 6,383 bp long with an average G+C content of 64.3%. A BLASTN search (2) revealed that the nucleotide sequence of this 6.4 kb area overall has a high level of identity to contiguous DNA regions in the following four microbial genomes listed below (GenBank accession number shown in parenthesis): *Burkholderia fungorum* (NZ_AAAC01000104), nucleotide identity 99-100%; *Xylella fastidiosa* (AE004000), nucleotide identity 84-88%; *Ralstonia metallidurans* (NZ_AAAI01000352), nucleotide identity 84-91%; *Xanthomonas axonopodis* (AE011858), nucleotide identity 83-86%; and to two *Pseudomonas aeruginosa* genomic islands: PAGI-2 (AF440523), nucleotide identity 84-90%, and PAGI-3 (AF440524), nucleotide identity, 73-80%. Protein coding sequences of eight major open reading frames identified in pTCB207 were assigned *orf1*
through 8 (Table 4, Fig. 5). The amino acid sequences predicted from the 7 ORFs also matched quite well to the predicted ORFs in the genomes mentioned above (Table 4). However, most of predicted proteins carried no known or hypothetical functions. In addition to the ORFs in the four genome sequences significant similarities (28 to 42% amino acid identity) were found between the ORFs predicted for this part of the clc island and ORFs from *Salmonella enterica* (AL513382) and *Pseudomonas fluorescens* (NZ_AABA01000162) (Table 4). Again, however, the proteins predicted from *orf2*, *orf4*, *orf5*, *orf6* and *orf7* matched solely hypothetical proteins of unknown function and carried no recognizable conserved domains. On the other hand, proteins predicted from *orf1* and *orf3* beared conserved domains of the ParA and ParB protein families, respectively (Table 4). Given that three up-regulating <TET-1> transposon insertions were mapped in *orf3* but two other neutral insertions as well (i.e., pTCB207-10 and –48, Fig. 5), this makes it unlikely that this gene (*orf3*) is coding for a negative regulator of the integrase. Rather, it suggests that the position of the three up-regulating <TET-1> transposon insertions influenced transcription of a down-stream located gene (like *orf4*) or, alternatively, that the product of the *orf3* gene influences the expression of yet another regulatory protein acting on the P_int promoter. Another ORF overlaps with the *orf3* sequence but in the other direction. This ORF encompasses all three upregulating but not the neutral <TET-1> insertions. Until the proper nature of this polypeptide is investigated in more detail, we can only speculate as to why this difference in mutant phenotype occurred.

**Repressor and activator binding sites.** To further investigate the nature of the ‘up’ and ‘down’ transposon insertions, GFP expression was assayed in two B13 derivative strains with the P_int promoter deletions fused to gfp in the presence of pTCB207, pTCB207-41 (with the ‘down’ mutation) or pTCB207-6 (with the ‘up’ mutation). Like in B13-jim1, introducing the plasmid pTCB207 into B13-jim2 (Fig. 3) resulted in an enhanced GFP production, i.e., the 95% percentile value increased with 43 and the top 5% mean value with 38 AGV units (Table 3). Interruption of the putative activator, *orf7*, weakened this stimulation (Table 3). Perhaps unexpected, but introduction of the plasmid pTCB207-6 into B13-jim2 resulted in an ‘up’ phenotype of GFP expression, and even in B13-jim6 an ‘up’ phenotype was observed (Table 3). This argued against a direct interaction between the product of *orf3* and the promoter area upstream of primer jim2 (Fig. 3), but rather points to the involvement of *orf3* in regulating expression of *orf7*.
**TABLE 4.** Location, proposed function and homologies of peptides predicted from ORFs within the 6.4 kbp EcoRI fragment at the left end of the clc genomic island.

<table>
<thead>
<tr>
<th>ORF</th>
<th>ORF location&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Length, nucleotides/ amino acids</th>
<th>Proposed function</th>
<th>Percentage of amino acids identity to ORFs in &lt;br&gt;Bfu&lt;sup&gt;b&lt;/sup&gt; Xfa Rme Xax Pagi-2 Pagi-3 Sen Pfl</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>30-905</td>
<td>876 / 291</td>
<td>Conserved hypothetical protein, similar to members of ParA family&lt;sup&gt;d&lt;/sup&gt;</td>
<td>100 91 93 76 93 71 39 -</td>
</tr>
<tr>
<td>ORF2</td>
<td>889-1146</td>
<td>258 / 85</td>
<td>Conserved hypothetical protein</td>
<td>76 - - - - - -</td>
</tr>
<tr>
<td>ORF3</td>
<td>1139-2791</td>
<td>1653 / 550</td>
<td>Conserved hypothetical protein, similar to members of ParB family&lt;sup&gt;e&lt;/sup&gt;</td>
<td>100 76, 91&lt;sup&gt;f&lt;/sup&gt; 81 58 86 61 42 42</td>
</tr>
<tr>
<td>ORF4</td>
<td>2807-3367</td>
<td>561 / 186</td>
<td>Conserved hypothetical protein</td>
<td>100 93 90 73 90 68 34 41</td>
</tr>
<tr>
<td>ORF5</td>
<td>3243-4616</td>
<td>1374 / 457</td>
<td>Conserved hypothetical protein</td>
<td>100 80 75 62 82 63 31 -</td>
</tr>
<tr>
<td>ORF6</td>
<td>4947-5726</td>
<td>780 / 259</td>
<td>Conserved hypothetical protein</td>
<td>100 91 86 77 90 80 28 52</td>
</tr>
<tr>
<td>ORF7</td>
<td>5723-6250</td>
<td>528 / 175</td>
<td>Conserved hypothetical protein, activator of the integrase</td>
<td>100 85 86 65 86 67 - -</td>
</tr>
<tr>
<td>ORF8</td>
<td>3118-2438</td>
<td>648 / 215</td>
<td>Hypothetical protein</td>
<td>- - - - - - - -</td>
</tr>
</tbody>
</table>

<sup>a</sup> First and last nucleotides as in the sequence deposited at GenBank under Accession number XXX.

<sup>b</sup> Bfu, Burkholderia fungorum, unfinished genome (NZ_AAAC01000104); Xfa, Xylella fastidiosa, complete genome (AE004000); Rme, Ralstonia metallidurans, unfinished genome (NZ_AAAB01000352); Pagi-2, Pseudomonas aeruginosa genomic island PAGI-2 (AF440523); Pagi-3, P. aeruginosa genomic island PAGI-3 (AF440524); Xax, Xanthomonas axonopodis, complete genome (AE011858); Sen, Salmonella enterica, complete genome (AL513382); Pfl, P. fluorescens, unfinished genome (NZ_AABA01000162).

<sup>c</sup> Similarity is not detected with BLASTP (1).

<sup>d</sup> Predicted polypeptide region between amino acid residues 104 and 194 aligns with the protein domain conserved among ParA-type ATPases (position 22-106) [pfam0991, E value 0.007].

<sup>e</sup> Predicted polypeptide region between amino acid residues 39 and 102 aligns with ParBc, ParB-like nuclease domain (position 3-93) [pfam02195; E value 8E-05].

<sup>f</sup> ORF3 corresponds to two consecutive ORFs in X. fastidiosa (XF1784 and XF1783).
DISCUSSION

Taken together, the results obtained here were largely consistent with the hypothesis that expression of P\textsubscript{int} is subject to positive and negative transcription control. All data so far confirm that the product of the \textit{orf7} gene would encode a transcriptional activator. Mutations in \textit{orf7} abolished the activation effect, and deletion of the promoter region to beyond jim6 abolished activation potential as well. Complication to this scenario is that the transcription start would take place in the same region as the supposed binding site of the ORF7 factor. RT-PCR mapping of the mRNA transcript strongly suggested the promoter to be upstream of the jim6 region. The finding that the jim6-\textit{gfp} fusion construct still displayed inducible GFP expression would indicate that the ORF7 activator could still bind to the DNA, but when the native promoter was deleted perhaps a weak RNA polymerase binding site was created by the fusion sequence.

The results of transposon mutagenesis were not so clear for the proposed repressor function. A repressor function was postulated on the basis of increased GFP expression when deleting part of the promoter/operator region (as in the jim2-\textit{gfp} fusion). However, although an ‘up’ phenotype was observed among the transposon mutants in plasmid pTCB207, it is unlikely that the \textit{orf3} gene codes for the actual repressor. Firstly, because two other transposon insertions in the same ORF had no effect on GFP expression levels. Secondly, because the same transposon mutants that caused the ‘up’ phenotype in B13-jim1, also caused an ‘up’ phenotype in B13-jim2, in which the presumed repressor binding site was removed. Rather, this suggests that the \textit{orf3} product indirectly affects \textit{intB13} expression, either by repressing the proposed activator \textit{orf7} or by causing a polar effect on downstream located genes, such as \textit{orf4}, in which no independent transposon insertion was obtained. Therefore, until a real repressor protein is found binding to the \textit{intB13} operator, it might be that a DNA structure itself or a small RNA exert the repressive effect. Confirmation of the roles of the \textit{orf3}, \textit{orf7} and \textit{orf4} gene products in \textit{intB13} expression will await protein-DNA studies.

Although we could not firmly establish the role of the ORF3 peptide, its homology to the ParB family of transcriptional repressors was significant. Upstream of \textit{orf3} and in the same direction we found an ORF, named \textit{orf1}, with relationships to ParA-type proteins. The \textit{parA parB} pair of genes is a well-known couple of chromosome- and plasmid-partitioning determinants but have other roles as well (17). For example, the \textit{korA/IncC-korB} pair in the central control operon of IncP-1 plasmids (41) determines plasmid partitioning, but also
exerts global transcription control over plasmid replication and conjugative transfer (6, 27, 28, 65). A ParB-like transcription regulator was also described in *Shigella* spp. controlling not partitioning but pathogenicity (7, 61). This may indicate that the *orf3* gene product of the *clc* genomic island is involved in regulatory aspects, perhaps, also of the conjugation process.

It appeared that the gene cluster on pTCB207 involved in regulation of the integrase of the *clc* island is well conserved also in other bacterial genomes (Table 4). At least four of these clusters occur on genomic islands, i.e. PAGI-2 and PAGI-3 in *P. aeruginosa* (34); one in *R. metallidurans* CH34 (34); and one in *X. fastidiosa* (59). All these elements carry also integrases strongly related to IntB13. Since the overall DNA sequences contained by the genomic islands of *X. fastidiosa* and *P. aeruginosa* is very different (34), this conservation of the left end region may be especially important for regulation of the integrase, excision or conjugative transfer.

It could be firmly established that transcription of the integrase *intB13* in the *clc* genomic island occurs from two alternative promoters, P*$_{circ}$* and P*$_{int}$*, depending on the element’s state in the cell. In the closed circular form *intB13* is transcribed from the P*$_{circ}$* promoter. This promoter directs transcription from the left end through the junction into the right end and the integrase gene. In the integrated *clc* element the arrangement of the integrase promoter is different, as a result of displacement of the left end promoter by tRNA*$_{Gly}$* gene. The 5’ end of the *intB13* transcript from the integrated form could be confined to a region between primer 2.6 and the 5’-half of jim6 (Fig. 3). There are conserved –35- (TTGAAA) and -10-like (TTTTTTT) boxes in this region, 17 base pairs apart, which, however, seem to lay too close to the start of the *intB13* transcript identified by RT-PCR (e.g., the TTTTTTT box lays completely within the jim6 primer region). It is not exceptional that conserved–10 and –35 elements are absent, since several other integrase promoters lack one or both of them. For example, the phage λ integrase promoter P$_i$ has only a –35-like element (12), whereas in phage 186 the P$_c$ promoter for immunity repressor and the integrase only has a –10-like element (53). Activation of these promoters is achieved by recruitment of RNA polymerase with CII transcriptional activators (24, 53). Similarly, the promoter of the *repSA-xis-int* operon in the pSAM2 integrative element in *Streptomyces* is devoid of both –10 and –35 boxes but requires for transcription an activator Pra (51, 52).

Our hypothesis for dual control of *intB13* transcription is based on a more general phenomenon occurring among temperate bacteriophages. For example, the alternative λ integrase promoter P$_L$ is under major control of the CI repressor and the Cro activator (16). Expression from P$_L$ is linked to physiology of the host via a complex network of direct and
indirect interactions of CII key regulator with CIII and with host-encoded functions (3, 30). In phage P2, the P_c promoter which controls expression of the immunity repressor C and the integrase, overlaps with the face-to-face P_e promoter for the early genes and Cox. Immunity repressor C both negatively and positively regulates its own promoter P_c (50) and represses P_c. Cox act as repressors for transcription from P_e (49), and the integrase expression depends on the ratio between C and Cox. The balance can be changed by intervention of the anti-repressor protein from the satellite phage P4 (15, 37). Regulatory system as in P2 also works in phage W (36). In some cases cascade control mechanisms are used for integrase expression. In the already mentioned integrative element pSAM2 integrase expression is activated by Pra, but pra transcription itself is inhibited by the global regulator KorSA (51, 52).

What is the significance of having two functionally distinct promoters for the integrase? The clc island is self-transmissible and spreads among bacteria via conjugation. Its establishment in the new host relies on site-specific integration of the circular intermediate, which would require a sufficient amount of the integrase. This can be at least partially ensured by having integrase expression in the circular form under control of strong constitutive promoter P_circ. Similar strategies occur in phage P4 (42) and in high pathogenicity island of Yersinia (43). Both P4 and the Yersinia genomic island form circular molecules when excised placing their integrases under control of constitutive promoters. After integration, intB13 is placed under control of the alternative promoter, P_int by physical separation of the P_circ promoter and the integrase gene. This is also similar for Yersinia genomic island and for the Streptomyces bacteriophage φC31 (33). What exactly causes the integrated P_int promoter to become activated is not yet very clear, although we observed that GFP expression was highest in stationary phase cultures. Perhaps induction of the integrase of the clc genomic island is part of a cascade aimed to rescue the island to another recipient when environmental conditions turn bad. Unraveling the regulatory network that links the host physiology and the activity of genomic island will broaden our knowledge of mechanisms underlying spread and evolution of a variety of integrative and conjugative elements.
EXPERIMENTAL PROCEDURES

**Strains and plasmids.** *Escherichia coli* DH5α (Gibco BRL, Life Technologies) was used routinely for plasmid propagation and transformation. *E. coli* HB101 (Promega) was used for the propagation of the helper plasmid pRK2013 (13), *E. coli* CC118λpir (32) was used to maintain plasmid pCK218 (32) and its derivatives. *E. coli* S17-1 (55) was used as donor for plasmids derived from pKT230 (5) *Pseudomonas* sp. strain B13 (14) is the original host of the *clc* element.

**Promoter reporter construction.** The approach used to construct integrase promoter reporters was as follows. Briefly, six DNA fragments in the range from 70 to 274 bp covering the integrase promoter region at *attR* were amplified by using the PCR with plasmid pRR108 (46) (see Fig. 3). Primers for PCR amplification were designed such as to introduce *Sph*I, *Pst*I or *Bam*HI restriction sites (Table 1) in order to facilitate subsequent cloning of the amplified fragments to a promoterless *gfp* gene. The *gfp* gene used here coded for the F64L,S65T enhanced green fluorescent mutant protein (39) and was subcloned into plasmid pJAMA23 (29). Inserts covering parts of the *intB13* upstream region were sequenced to detect possible changes introduced by PCR amplification. A 480-bp *Eco*RI fragment containing the region upstream of the *intB13* gene in the circular form of the *clc* element was retrieved from plasmid pRR146 (46) and cloned into pJAMA23 in front of the *gfp* gene. All resulting *intB13* promoter-*gfp* fusions were retrieved as single *Not*I fragments and cloned into the mini-Tn5 delivery plasmid pCK218 (32), thereby replacing the *luxAB* genes. Subsequent random insertion of the *intB13* promoter-*gfp* fragments into the genome of *Pseudomonas* sp. strain B13 was achieved through mobilization of the pCK218 derivative plasmids from *E. coli* CC118λpir by helper plasmid pRK2013 in a triparental filter mating as described by Herrero et al. (23). The presence and sizes of the *intB13* promoter-*gfp* fragment and the absence of the plasmid backbone in B13 transconjugants were verified by antibiotic resistance profiling and by using the PCR with two construct-specific primers (i.e., Ter-Fw and gfp-R, Table 1). Clones with correct insertions were assigned *Pseudomonas* sp. strain B13 (jim1 through 9) (see Fig. 3 for details).

**Media and growth conditions.** Luria–Bertani broth (LB) was routinely used for growing *E. coli* and *Pseudomonas* strains. As a defined mineral medium (MM) the type 21C mineral medium (18) was used, supplemented either with 10 mM CBA or with 10 mM fructose. When required 50 μg of ampicillin, kanamycin, rifampin, streptomycin and/or 5 μg
tetracycline per ml were added. Strains of *Pseudomonas* were grown at 30°C, those of *E.coli* at 37°C.

**Cultivation of the reporter strains.** Constructed reporter strains of *Pseudomonas* sp. B13 were stored in 20% glycerol stocks at –80°C and whenever necessary plated on MM agar with the appropriate antibiotics and 5 mM CBA as sole carbon source. To obtain a liquid culture for microscopy a single colony was inoculated into 5 ml of MM supplemented with 10 mM CBA or fructose in a 35 ml glass tube and shaken at 200 rpm. Typically, exponentially growing cultures were obtained within 16 h, and after between 24 and 48 h of cultivation the cultures reached stationary phase.

**Fluorescence microscopy.** GFP fluorescence intensities of individual cells of *Pseudomonas* sp. strain B13 derivatives were examined with an Olympus BX50 epifluorescence microscope. Cells from liquid cultures were tenfold concentrated by centrifugation and resuspension, applied to glass slides, covered with a cover slip and directly studied under the microscope. Cells attached to the cover slip were focused. Images of at least 200 cells per field were taken with a cooled CCD camera (Photometrics SenSys:1401E, Roper Scientific Inc., USA), a 100 x / 1.30 oil immersion lens (UPlanF1, Olympus, Japan) and an exposure time of 300 msec. The filter used for GFP fluorescence was HQ-EGFP (F41-017, AF Analysentechnik, Tübingen, Germany) with emission filter HQ 525/50, beamsplitter Q 495 LP and excitation filter HQ 470/40. Both phase-contrast and GFP fluorescence images of 6 to 12 randomly chosen fields per sample were recorded with the ACQUIRE function of the METAMORPH Software (Version 4.6, Universal Imaging Corp.). Samples to be directly compared were analyzed on the same day to avoid possible UV light source intensity variations.

**Image analysis.** Quantification of fluorescence intensities of individual cells in an image was performed in an automatic subroutine within the METAMORPH software. To account for background variation, a low pass function was generated, which was subtracted from the original image. Then, a lower threshold was defined to remove background and to allow automatic object identification. The average gray value (AGV), the total gray value of all pixels per cell and the total area were recorded automatically for each individual object (cell) as well as summed up for all objects of an image. Average gray values were then pooled together for all images of each particular culture sample or condition. In each culture sample a minimum of 2000 cells was analyzed.

**Population analysis.** Differences in cellular average gray values among samples were analyzed by determining the distribution of AGVs for all cells in a population.
purpose, the AGVs of all cells were sorted, ranked and plotted against their position in the cumulative distribution curve (i.e., the ranking number divided by the total number of cells in the population, multiplied by 100). A statistical subroutine of the program R (25) (http://cran.r-project.org) was written in order to analyze differences in the cumulative distribution curves based on non-distribution type functions. The most prominent parameters for comparing statistical differences among samples were found to be the AGV (or GFP fluorescence value) corresponding to the 95% percentile of the sample population and the arithmetic mean of the cellular AGV values for those 5% of the population with the highest single AGVs (indicated as ‘Top 5% Mean’). Intervals for 95% confidence were calculated for each of the derived 95% percentile fluorescence values based on 200 bootstrapping cycles.

**Transcription initiation start mapping using reverse transcription.** Total cellular RNA was isolated using the acid phenol extraction protocol (1) followed by DNase I (Amersham) treatment, phenol/chloroform purification and precipitation with 70% ethanol and 0.3 M sodium acetate (pH 5.2). The reverse transcriptase (RT) reaction was done with 1 µg total RNA by using the Enhanced Avian RT First Strand Synthesis kit (Sigma) according to the guidelines of the supplier (Sigma). The reverse transcriptase reaction was initiated from primer SV004-BamHI-R (Table 1), which was labelled at its 5’ end with the fluorescent dye IRD-800 (purchased from MWG Biotech, Münich, Germany). The resulting RT products were directly analyzed on an automated sequencer (model 4200IR2; LI-COR Inc., Lincoln, Nebr.). Parallel reference sequencing reactions were produced with the same primer on plasmid pJAMA23-jim4 by using the Thermo Sequenase fluorescence labeled primer cycle kit with 7-deaza-dGTP (Amersham). To approximately map the transcription initiation start in those RNA samples with low abundance of intB13 specific RNA transcripts we carried out RT-PCR. cDNA obtained by the RT reaction on total RNA with primer int271-R (Table 1) was amplified by PCR as follows. 50 µl RT reaction mixture was treated for 1 h with 1 unit of DNAse-free RNAse (Boehringer Mannheim) and 5 µl used in the PCR with either one of the forward primers jim1, jim2, jim2.6, jim3, jim4.1, jim6, jim7 and jim9 (Fig. 3, Table 1) in combination with the reverse primer int271-R. Each PCR reaction mixture contained Sigma PCR buffer, 250 µM of each deoxynucleoside triphosphate (Sigma), 1 unit of TAQ polymerase (Sigma), and 1 pM of each primer in 50 µl final volume. The following cycling parameters were implemented: 2 min hot-start at 95°C followed by 35 cycles of 30 s denaturation at 93.5°C, 30 s of annealing at 50°C and 45 s of extension at 72°C, with 4 min of final extension at 72°C. As negative control for the RT-PCR reaction total RNA in RT
reaction mix but without reverse transcriptase enzyme was used. As positive control, purified genomic DNA of strain B13 (jim1) was amplified in the PCR, without RT-step.

**Cloning and mapping of genes involved in P_int regulation.** Two EcoRI fragments of 4.2 and 6.4 kb size, originating from the left end of the clc element (see Fig. 5) were cloned from the cosmid 3G3 (45) into pUC28, which resulted in plasmids pTCB172 and pTCB177, respectively. Plasmids pTCB172 and 177 were cleaved with HindIII and ligated with pKT230 (5) giving rise to pTCB208 and pTCB207, respectively. Both plasmids contained stable co-replicons and were maintained in E.coli S17-1 under ampicillin and streptomycin selection. To assess if these two DNA fragments interfered with expression of the P_int-gfp fusion, plasmids pTCB207 and pTCB208 were transferred by filter mating from E. coli S17-1 to Pseudomonas sp. B13-jim1. Expression of GFP after 7 days of growth in liquid MM with 10 mM CBA in Pseudomonas sp. strain B13-jim1 (pCBA207) or (pCBA208) was compared to that in strain B13-jim1 without any plasmid and strain B13-jim1 (pKT230). Further functional mapping was performed by mutagenizing purified plasmid pTCB207 in vitro by using the EZ::Tn<sup>TM</sup> <TET-1> Insertion Kit (Epicentre Technologies Inc.) according to the manufacturer’s recommendations. After transformation into E. coli S17-1 mutant plasmids carrying the randomly inserted tetracycline resistance transposon <TET-1> were isolated and screened with EcoRI restriction analysis for insertions in the 6.4 kb EcoRI fragment. DNA sequences flanking the <TET-1> insertion in twelve selected plasmids were determined by DNA sequencing using primers internal to <TET-1> facing outwards (Epicentre Technologies Inc.). These twelve plasmids, assigned pTCB207-6, -7, -10, -41, -46, -48, -52, -53, -62, -64, -66 and –68 (Fig. 5) were subsequently introduced into Pseudomonas sp. strain B13-jim1 for testing the effect of <TET-1> insertions on GFP production. Plasmid pTCB207, pTCB207-6 and –41 were transferred to B13-jim2, and pTCB207-6 to B13-jim6 to assess their influence on transcription from the jim2 and jim6 shortened integrase promoters.

**DNA sequencing and analysis.** DNA sequencing was performed on double stranded DNA templates with a Thermo Sequenase cycle sequencing kit with 7-deaza-dGTP (Amersham). Sequencing reactions were analysed on an automated DNA sequencer model 4200 IR<sup>2</sup> (LI-COR Inc., Lincoln, NE, USA). Primers for sequencing were labeled with fluorescent dyes IRD-800 and IRD-700 at 5’ end and were purchased from MWG Biotech (Ebersberg, Germany). Sequence assembly and computer analysis was done with DNASTAR software (DNASTAR Inc., Madison, WI, USA). The strategy to sequence the insert of pTCB207 consisted of constructing overlapping subclones plus sequencing outwards from <TET-1> insertions. Sequences were aligned using the program SEQMAN within the
DNASTAR package. Multiple (5 to 19) single sequences read in both directions covered the complete region. A few conflict positions in the program-generated consensus sequence were manually corrected by checking raw sequencing data files. The resulting DNA sequence is deposited in GenBank under Accession number xxxx. Open reading frames were identified with the DNASTAR application MAPDRAW and compared with sequences in GenBank and EMBL using the BLAST search tools (2).
REFERENCES


54. Shoemaker, N. B., G. R. Wang, and A. A. Salyers. 1996. The *Bacteroides* mobilizable insertion element, NBU1, integrates into the 3' end of a Leu-tRNA gene and has an integrase that is a member of the lambda integrase family. J. Bacteriol. 178:3594-3600.


A case for directed evolution:
excision and transfer of the clc genomic island
in Pseudomonas sp. strain B13 is triggered by starvation
and stimulated by the environmental pollutant,
3-chlorobenzoic acid
SUMMARY

*Pseudomonas* sp. strain B13 carries a 105 kb conjugative genomic island (the *clc* element) which encodes catabolism of 3-chlorobenzoic acid (CBA). The *clc* element normally resides in two copies in the B13 chromosome, but can excise and transfer itself to a new recipient, where reintegration takes place into a specific target site. Central for the excision process and for site-specific integration in the new bacterial host’s chromosome is the enzymatic activity of the integrase, IntB13, encoded near one end of the *clc* element. By creating transcription fusions between the *intB13* promoter and the *gfp* gene, by analyzing the formation of the excised form of the *clc* element and by quantitative conjugation experiments we show that integrase expression in strain B13 is inducible and occurring primarily under stationary phase conditions. Strangely, integrase expression was significantly stimulated by growing cultures on CBA. High cell density, heat shock, osmotic shock, UV irradiation, or treatment with alcohol did not stimulate integrase expression. To our knowledge this is the first report where a chlorinated compound stimulated horizontal transfer of the genes encoding its very metabolism.
INTRODUCTION

Bacteria can adapt to changing environmental conditions by a variety of genetic mechanisms. Some of these mechanisms involve acquisition of foreign DNA through mobile DNA elements like plasmids, transposons or genomic islands. These processes of foreign DNA acquisition, also named horizontal gene transfer (HGT) have been focus of extensive research because of their general importance for microbial evolution, for the formation of catabolic pathways and of antibiotic resistance in particular (14, 36) and for pathogenicity determinants (15, 19). Little attention was given until now to the possibility of regulation of HGT processes by signalling pathways, effector compounds and/or environmental stimuli. The only major exceptions we are aware of, are i) the transfer of the opine catabolism encoding plasmids in Agrobacterium which is regulated by both catabolism of opines and quorum sensing (10-12, 26), ii) regulation of transfer of the Bacteroides conjugative transposons by tetracycline (5, 30) and iii) the stimulation of conjugation competence formation by peptide pheromones in Enterococcus faecalis (reviewed in (8)). The possibility for regulation of HGT by environmental stimuli is intriguing, because it implicates that conditions stimulating or repressing certain types of HGT may be created due to human activities. Furthermore, a direct link between environmental signals and HGT points to the existence of mechanisms of directed evolution.

In the present study we describe a possible direct link between the transfer of the clc genomic island of Pseudomonas sp. strain B13 (9), which carries the genes for 3-chlorobenzoic acid (CBA) degradation, and the presence of CBA itself. The clc genomic island (also referred to as clc element (37)) is a mobile DNA element of 105 kb, which in strain B13 predominantly resides in the chromosome as two copies, integrated in genes for glycine tRNA (27, 28). The clc element is an interesting model system for the life style of genomic islands, since it is self-transmissible to a number of different Gram negative γ- and β-Proteobacteria, like Pseudomonas spp., Ralstonia spp., Alcaligenes spp. and Burkholderia spp. (27, 29, 35, 40). Self-transfer starts when the clc genomic island excises from the chromosome and forms a circular intermediate, which is then thought to travel to a new recipient and reintegrate into the chromosome. In all recipient strains analyzed so far, the clc element had always integrated site-specifically into one or more copies of the gene for glycine tRNA (27). The process of integration could be attributed to the intB13 integrase gene (Fig. 1A) proximal to one end of the clc element (28). Like in analogous systems, such as the
conjugative transposons Tn916 (32) and CTnDOT (6, 7), the activity of the IntB13 integrase is probably also catalyzing excision of the genomic island from the chromosome.

In this work, we analyzed the effect of CBA on the induction of the \textit{intB13} integrase and on transfer of the \textit{clc} element under different growth conditions. The analyses were based on measuring single-cell GFP fluorescence of cells containing fusions of the \textit{intB13} promoter with the gene for the green fluorescent protein (GFP), by measuring the appearance of the excised circular form of the \textit{clc} element and the transfer frequencies of the phenotype to metabolize CBA in bacterial matings.
MATERIALS AND METHODS

Strains and culture conditions. *E. coli* strains DH5α, HB101 (31) and CC118 λpir (20) were used and grown as previously described (18). *Pseudomonas* sp. strain B13 (9) is the original host and donor of the *clc* element. *Pseudomonas putida* strain UWC1 (23) (kindly provided by Carole Newberry, Cardiff University, Wales, UK) was used as recipients for the *clc* element. Luria Bertani broth (LB) was routinely used for growing *E. coli* and *Pseudomonas* strains. As a defined mineral medium (MM) the type 21C mineral medium (13) was used, supplemented with 10 mM CBA or fructose. When required 50 µg/ml of ampicillin, kanamycin and/or rifampin were added. Strains of *Pseudomonas* were grown at 30°C, those of *E. coli* at 37°C.

Promoter reporter constructions. Integrase promoter reporters were constructed by amplifying a 282 bp fragment of the P_\text{int} promoter (Fig. 1) by PCR and fusing this fragment via several steps to a promoterless *gfp* gene, coding for the F64L,S65T enhanced green fluorescent mutant protein (24). Similarly, the P_\text{circ} promoter region corresponding to the region upstream of the *intB13* gene in the circular form of the *clc* element, was cloned from pRR146 (28) as a 480 bp BamHI fragment and fused to the *gfp* gene. The resulting P_\text{int} and P_\text{circ}-\text{gfp} fusions were further cloned as *NotI* fragments into the mini-Tn5 delivery vector pCK218 (20) and randomly inserted into the genome of *Pseudomonas* sp. strain B13 by mobilization and subsequent transposition in triparental filter mating (16). Strains with the correct insertion were verified by PCR, antibiotic resistance profiling and Southern hybridization and were named *Pseudomonas* sp. strain B13-jim1 (P_\text{int}-\text{gfp}) and -jim4 (P_\text{circ}-\text{gfp}).

Induction experiments. *Pseudomonas* sp. B13-jim1 was grown in batch in 200 ml minimal medium (MM) containing either 10 mM CBA or 10 mM fructose at 30°C with 200 rpm shaking. Samples for optical density measurements (OD, determined at a wavelength of 600 nm), GFP expression and for DNA extraction were taken at 1 to 6 h intervals or whenever appropriate. Under continuous cultivation, strain B13-jim1 was grown in a 500 ml chemostat with a working volume of 200 ml, fed with MM supplemented with either 0.1 mM or 10 mM CBA, or 10 mM fructose. The chemostat was operated at a constant dilution rate of 0.05 h^{-1}. OD and GFP expression were analyzed daily on 2 ml samples taken aseptically from the reactor. To study effects of starvation, 1 ml aliquots of the chemostat cultures were incubated in 15 ml closed polypropylene tubes shaken horizontally at 200 rpm at 30°C and
FIG. 1. A) Schematic presentation of the two forms of the clc genomic island and of the reactions catalyzed by the IntB13 integrase. During integration the 18-bp 3' end (depicted by open triangles) of the target glycine tRNA gene (glyV) is replaced by an identical 18-bp sequence (filled triangles) originating from the right (R) end of the clc element. Excision results in a closed junction between left and right ends of the element (depicted L/R). Promoter regions of the integrase gene (intB13) of the integrated (P\text{int}) and the excised circular form (P\text{circ}) of the clc element are depicted by thin arrows showing the direction of transcription (not to proportion). B) P\text{circ}-gfp and P\text{circ}-gfp fusions present in strains B13-jim1 and B13-jim4, respectively. Solid vertical bars correspond to the I- and O-ends of the Tn5 delivery system.

A limited number of typical stress factors were tested on 2 ml aliquots of strain B13-jim1 withdrawn from continuous cultures fed with 10 mM fructose. To create a heat shock, aliquots were incubated for 5 minutes at 42°C. For UV treatment the test culture was dispensed into a 90-mm-diameter plastic petri dish and exposed to 2 J/cm² of UV radiation at 254 nm in UV Stratalinker 1800 (Stratagene, La Jolla, CA, U.S.A.). High osmolarity conditions were achieved by adding sodium chloride to the culture at final concentrations of 200 and 500 mM. Ethanol stress was performed by incubating cell aliquots with 10 % ethanol (vol/vol). After UV treatment and heat shock and during alcohol and high salt treatment, the samples were incubated on a shaker at 180 rpm and 30°C for 3-4 h, then divided into two portions: one for GFP expression measurements and another for total DNA isolation.
Transfer of the clc genomic island on membrane filters. To determine if the growth substrate would influence the frequency of clc element transfer, membrane filter matings were set up between exponentially grown cultures of *Pseudomonas* sp. strain B13-jim1 as a donor and *P. putida* strain UWC1 as a recipient. Volumes of both donor and recipient cultures in the mating were chosen such as to give approximately $10^8$ cells per filter each. Cells were collected by centrifugation at $3,000 \times g$ for 5 min, resuspended gently in 50 µl of MM and transferred to cellulose nitrate filters (0.45 µm pore size, 25 mm diameter, Sartorius AG, Goettingen, Germany) which were placed on MM agar plates with 1 mM CBA, with 1 mM fructose, or without C-source. Directly after placement and after 4, 24, 48 and 72 hours of incubation the cells were washed from the filters with 0.5 ml MM by vigorous vortexing, serially diluted and plated on MM agar supplemented with 5 mM CBA and 50 µg/ml rifampin to select for transconjugants. Colonies visible by naked eye were scored after 4 days of incubation. The number of donor and recipient cells at the end of mating was determined from serial dilutions on LB plates supplemented with either 50 µg/ml kanamycin or 50 µg/ml rifampin, respectively. The transfer frequency was calculated as the number of transconjugants per donor cell in the cell suspension washed from the filter. The transfer rate for every 24 h incubation interval was calculated as the net increase in the number of transconjugants per donor cell per hour during the corresponding incubation period.

Detection and quantification of the excised circular form of the clc element. Isolation of total genomic DNA was performed using the cetyltrimethylammoniumbromide (CTAB) method (2). Genomic DNA samples were digested with *Eco*RI, size-separated in 0.8% agarose gel, Southern blotted to Hybond-XL membrane (Amersham Bioscience) and hybridized with radioactively labeled DNA probes according to standard protocols (2). To detect the clc element in its integrated and excised circular form, a 1.7-kb *SalI*-NsiI fragment with the left-end of the clc element from plasmid pRR104 (27) was used in hybridizations. Band intensities of the circular form and the two chromosomally integrated clc element copies were estimated from autoradiograms by scanning the X-ray films as described previously (3). To verify the physical presence of covalently closed circular molecules of the clc element, the plasmid enriched DNA fraction was extracted in parallel with the total genomic DNA following the alkaline lysis protocol (4) and examined with Southern hybridization as above.

Fluorescence microscopy. GFP fluorescence intensities of individual cells were examined with an Olympus BX50 epifluorescence microscope. Images of at least 200 cells per field were taken with a cooled black-and-white CCD camera (Photometrics SenSys:1401E, Roper Scientific Inc., USA), a 100x /1.30 oil immersion lens (UPIanF1,
Olympus, Japan) and an exposure time of 300 msec. Digital imaging of GFP fluorescence and calculation of fluorescence intensities of each individual cell in an image were quantified in an automatic subroutine as described previously (18) and expressed as cellular average gray values (AGV).

**Population analysis.** Differences in cellular AGV among samples were analyzed by determining the distribution of AGVs for all cells in a population. For this purpose, the AGVs of all cells were sorted, ranked and plotted against their position in the cumulative distribution curve (i.e., the ranking number divided by the total number of cells in the population, multiplied by 100). A statistical subroutine of the program R (17) (http://cran.r-project.org) was written in order to analyze differences in the cumulative distribution curves based on distribution-free functions. The most prominent parameters for comparing statistical differences among samples were found to be the AGV (or GFP fluorescence value) corresponding to the 95% percentile of the sample population (Fig. 2A) and the arithmetic mean of the cellular AGV values for those 5% of the population with the highest single AGVs (indicated as ‘Top 5% MFV’). Intervals for 95% confidence were calculated for each of the derived 95% percentile fluorescence values based on 200 bootstrapping cycles.
RESULTS

**Distribution independent expression of the intB13 gene.** To determine activity of the intB13 gene of the clc genomic island in *Pseudomonas* sp. strain B13, we first relied on observing transcriptional activity from an extra copy of the upstream region of the intB13 gene (as in the integrated clc element) fused to a promoterless gfp gene and inserted at random into the *Pseudomonas* sp. strain B13 chromosome by means of Tn5 delivery (Fig. 1). In this case, the production of GFP in individual cells served as an indicator of the integrase promoter transcription activity.

Surprisingly, only a small proportion of a population of B13-jim1 cells containing the transcription fusion between the region upstream of the intB13 gene and the gfp showed GFP production (Fig. 2). The proportion of induced cells and the amount of accumulated GFP in induced cells of B13-jim1 was found to be dependent on the growth substrate and on the growth phase of the culture (see below), but remained small under all conditions tested. This extreme heterogeneity in GFP expression made it impossible to use common statistical descriptors such as an average fluorescence for the whole population with standard deviation as a measure for induction from the intB13 promoter. Therefore, we chose to use distribution-free statistical methods, such as the 95% percentile fluorescence value (Fig. 2A) and the mean fluorescence value (MFV) of the top 5% cells (Fig. 2A, inset) to compare expression from P_int under different conditions. To ensure that the observed gfp expression from the P_int-gfp fusion was not a peculiarity of GFP in strain B13 in general, we determined GFP expression from a P_circ-gfp fusion (Fig. 1) in single copy inserted on the chromosome by the same Tn5 delivery system (Fig. 1B). In this case, the distribution of GFP fluorescence intensities over the population could be described with a log normal distribution function (Fig. 2B). Furthermore, to ensure that the observed expression from the P_int-gfp fusion was not due to the chromosomal location of that particular insertion, we screened four independent insertions. All behaved the same with regard to single cell GFP fluorescence intensity and distribution levels in the population (data not shown).

**Induction of the integrase promoter under stationary phase conditions.** During the exponential phase of batch growth GFP production in cultures of B13-jim1 was very low both on fructose and CBA (Table 1). Transition to stationary phase significantly increased expression from the P_int-gfp fusion, although at most we observed 13% of the population with GFP intensities higher than 5 units above background (i.e., a grey value of 79). In stationary
FIG. 2. A) Distribution of GFP fluorescence intensities in a population of cells of *Pseudomonas* sp. strain B13-jim1 carrying the \( P_{\text{int}}-\text{gfp} \) fusion taken at different growth phases. Shown are cumulative distribution plots in order to visualize the concepts of 95% percentile and top 5% mean fluorescence value (MFV). The GFP fluorescence is given in average gray value units per cell calculated with METAMORPH. A fluorescence value of 79 corresponds to dark (non-fluorescent) cells and is an arbitrary value produced by the image recording system. The inset shows the range of fluorescence values for the whole population, whereas the main panel focuses on the range above 80%. Note the strong aberration from the Normal distribution (see B). Cells were grown in batch with 10 mM CBA as C-source. Samples were taken at 14 h (exponential growth phase) and 88 h (stationary phase). B) Cumulative distribution of the GFP fluorescence values in a population of cells of *Pseudomonas* sp. strain B13-jim4 carrying a \( P_{\text{circ}}-\text{gfp} \) fusion (see Fig.1). Each point indicates the fluorescence value of a single cell. Cells were grown in batch with 10 mM CBA as C-source, samples were taken at 14 h (curve a, exponential growth phase) and 40 h (curve b, stationary phase). In this case the cumulative distribution is represented very well by the Normal distribution (modelled in curve a’ and b’, respectively). Note the different scales of fluorescence in panels A and B. C) and D) Micrographs of typical population differences of strain B13-jim1 and of strain B13-jim4, respectively, taken in stationary phase under GFP illumination (left) and the corresponding image in phase-contrast (right).
TABLE 1. GFP fluorescence intensity from the P_int-gfp fusion in strain B13-jim1 grown in batch with fructose or 3-chlorobenzoic acid as C-source.

<table>
<thead>
<tr>
<th>Incubation time, h</th>
<th>10 mM fructose</th>
<th>10 mM 3-chlorobenzoate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GFP fluorescence intensity</td>
<td>GFP fluorescence intensity</td>
</tr>
<tr>
<td></td>
<td>95% percentile&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Top 5% MFV&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>14&lt;sup&gt;c&lt;/sup&gt;</td>
<td>79.6 (79.5-79.6)</td>
<td>80.4</td>
</tr>
<tr>
<td>40</td>
<td>81.8 (81.6-82.0)</td>
<td>86.8</td>
</tr>
<tr>
<td>64</td>
<td>84.0 (83.5-84.8)</td>
<td>92.4</td>
</tr>
<tr>
<td>88</td>
<td>83.3 (82.2-84.1)</td>
<td>91.7</td>
</tr>
</tbody>
</table>

<sup>a</sup>) GFP fluorescence intensity is given as METAMORPH cellular average gray value units of the 95% percentile of the population. Values within parentheses represent the 0.95 confidence intervals calculated with the R-software.

<sup>b</sup>) Values given are the arithmetic mean of the cellular average gray values for those 5% of the population with highest fluorescence intensities.

<sup>c</sup>) At 14 h both fructose and 3-chlorobenzoate cultures were growing exponentially and entered stationary phase after about 24 h incubation.

Phase both the proportion of GFP containing cells and their fluorescence levels were higher with CBA than with fructose as carbon source (Table 1).

Cells taken from continuously cultivated chemostats at a dilution rate of 0.05 h<sup>-1</sup> (which is close to stationary phase growth) displayed high levels of GFP production similar to stationary phase cells grown in batch culture. Likewise, fluorescence of the cells grown with CBA was higher compared to fructose (95% percentile fluorescence value corresponded to 100.8 and 84.2 AGV units respectively). When cells were taken out of the chemostat with 10 mM fructose and starved for carbon in subsequent batch incubation no further increase of the GFP fluorescence values occurred (Table 2). The relative proportion of induced cells up to 112 h of starvation also remained constant (about 5%). In contrast, B13-jim1 bacteria taken from CBA-fed chemostats increased both the proportion of induced cells (from 7% in chemostat at steady state to 13% after 64 h of starvation) and GFP production (from 100.8 units of population 95% AGV percentile in steady state to 105.6 after 64 h starvation) (Table 2). Spiking cells taken from the chemostat with 10 mM fructose, in batch with 1 mM or 5 mM CBA did not result in any further increase of the cells’ fluorescence during 4 h of incubation (Table 3). Batch and chemostat experiments were performed twice independently with similar results.
TABLE 2. GFP production from the P\textsubscript{int}gfp fusion in strain B13-jim1 under carbon limitation and starvation conditions

<table>
<thead>
<tr>
<th>Incubation time, h</th>
<th>Fructose\textsuperscript{a}</th>
<th>3-chlorobenzoate\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GFP fluorescence intensity\textsuperscript{b}</td>
<td>95% percentile</td>
</tr>
<tr>
<td>0\textsuperscript{c}</td>
<td>84.2 (82.4-85.9)\textsuperscript{d}</td>
<td>108.6</td>
</tr>
<tr>
<td>16</td>
<td>85.9 (84.5-88.7)</td>
<td>109.1</td>
</tr>
<tr>
<td>40</td>
<td>84.7 (84.0-86.9)</td>
<td>105.9</td>
</tr>
<tr>
<td>64</td>
<td>84.0 (83.8-84.4)</td>
<td>99.5</td>
</tr>
<tr>
<td>112</td>
<td>83.9 (83.3-85.4)</td>
<td>99.0</td>
</tr>
</tbody>
</table>

\textsuperscript{a}) Cultures were taken from chemostats continuously growing with 10 mM fructose or 10 mM 3-chlorobenzoate at dilution rate 0.05 h\textsuperscript{-1}.

\textsuperscript{b}) GFP fluorescence values are given as for Table 1.

\textsuperscript{c}) Time zero indicates the culture during steady state chemostat growth; subsequent time points refer to cells taken from the chemostat and incubated in batch without any carbon.

\textsuperscript{d}) For the time zero 95% percentile and top 5% MFV values present mathematical average of 3 samples taken from the chemostat daily. 95% confidence intervals shown for this time point were calculated using function CONFIDENCE within EXCEL.

**Formation of the circular form of the clc element.** To determine if measurements of expression from the integrase promoter by GFP fluorescence were a proper indicator for integrase activity, we measured the product of excision of the clc genomic island in strain B13-jim1 cultures in the same growth experiments as described above. As Fig. 3A shows, during exponential growth in batch with 10 mM CBA as C-source the amount of the circular form of the clc element remained undetectable (Fig. 3A, lane 1). However, shortly after the culture entered stationary phase, the circular form appeared (Fig. 3A, lane 2) and within the next 24 h its proportion increased to about 16% of the total hybridizable clc DNA (i.e., circular form plus chromosomal copies) according to densitometric analysis. This relative level of excised form remained more or less constant for the next 48 h (not shown).

In cells grown on fructose this value increased maximally to approximately 8%, or half the amount of circular form in cultures grown with CBA to the same growth phase (Fig. 3A, lanes 4 and 5). In cultures taken from the chemostat and incubated under conditions of carbon starvation, the proportion of the excised form compared to both chromosomal copies increased from about 4 % under steady state conditions with 10 mM CBA and a growth rate of 0.05 h\textsuperscript{-1} to 15% after 64 h of starvation (Fig. 3B).
TABLE 3. Effect of CBA supplement and various stresses on GFP production from the P<sub>m</sub>:<i>gfp</i> fusion in strain B13-jim1

<table>
<thead>
<tr>
<th>Treatment</th>
<th>GFP fluorescence&lt;sup&gt;b&lt;/sup&gt;</th>
<th>95% percentile</th>
<th>Top 5% MFV</th>
</tr>
</thead>
<tbody>
<tr>
<td>No treatment</td>
<td>83.4 (83.2-83.6)</td>
<td>86.3</td>
<td></td>
</tr>
<tr>
<td>1 mM CBA</td>
<td>83.4 (83.2-83.5)</td>
<td>85.8</td>
<td></td>
</tr>
<tr>
<td>5 mM CBA</td>
<td>83.4 (83.1-83.6)</td>
<td>85.8</td>
<td></td>
</tr>
<tr>
<td>No treatment</td>
<td>85.7 (85.5-86.0)</td>
<td>89.1</td>
<td></td>
</tr>
<tr>
<td>UV</td>
<td>85.1 (84.9-85.4)</td>
<td>88.3</td>
<td></td>
</tr>
<tr>
<td>42&lt;sup&gt;0&lt;/sup&gt;C, 5 min</td>
<td>83.7 (83.5-83.9)</td>
<td>85.7</td>
<td></td>
</tr>
<tr>
<td>0.2 M NaCl</td>
<td>84.5 (84.3-84.6)</td>
<td>87.3</td>
<td></td>
</tr>
<tr>
<td>0.5 M NaCl</td>
<td>86.7 (86.4-87.1)</td>
<td>89.4</td>
<td></td>
</tr>
<tr>
<td>10% ethanol</td>
<td>84.3 (84.1-84.6)</td>
<td>86.6</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>) Cultures were taken from steady state growing chemostat culture on 10 mM fructose. Two-milliliter aliquots were treated as described in Materials and Methods.  
<sup>b</sup>) GFP fluorescence intensity is given as for Table 1.

TABLE 4. Effect of the cell density on GFP production from P<sub>m</sub>:<i>gfp</i> fusion in the strain B13-jim1

<table>
<thead>
<tr>
<th>Cell culture</th>
<th>0.1 mM CBA (OD&lt;sub&gt;600&lt;/sub&gt;=0.01)</th>
<th>10 mM CBA (OD&lt;sub&gt;600&lt;/sub&gt;=0.83)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95% percentile</td>
<td>Top 5% MFV</td>
</tr>
<tr>
<td>Chem 6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>83.8 (83.7-84.2)</td>
<td>86.0</td>
</tr>
<tr>
<td>Chem 14&lt;sup&gt;d&lt;/sup&gt;</td>
<td>83.8 (83.5-84.1)</td>
<td>88.8</td>
</tr>
<tr>
<td>Starv 60 h&lt;sup&gt;b&lt;/sup&gt;</td>
<td>85.1 (83.8-87.4)</td>
<td>94.5</td>
</tr>
</tbody>
</table>

<sup>a</sup>) Cultures were taken from continuously growing chemostats on 0.1 or 10 mM 3-chlorobenzoate.  
<sup>b</sup>) Cells were taken from the chemostat on the 6<sup>th</sup> day and incubated in batch without any carbon for 60 h.  
<sup>c</sup>) GFP fluorescence intensity values are given as for Table 1. Note that absolute values of GFP fluorescence in this table can not be compared to those in the Table 2 due to slight changes in brightness of UV source.

Analysis of an enriched ‘plasmid’ DNA fraction from the cells starved for 64 h showed a much higher abundance of the 17-kb band compared to the total genomic DNA (Fig. 3B), which indicated that the hybridizing 17-kb EcoRI fragment resulted from the excised circularized <i>clc</i> element and not from tandemly repeated chromosomal copies.
FIG. 3. Appearance of the free circular form of the clc element in starving cells of *Pseudomonas* sp. strain B13-jim1 on Southern-blots of *Eco*RI digested genomic DNA hybridized against a probe for the left end of the element (see *Materials and Methods*). The two chromosomal copies of the element appear as two lower bands (sizes approximately 2 and 7 kb) whereas the uppermost band (ca. 17 kb) represents the excised circular form. A) Batch cultures. Lane 1, mid-exponential stage of growth (14 h) with 10 mM CBA; lane 2 and 3, same culture 3 h and 7 h after entering stationary phase, respectively; lane 4 and 5, 48 h old batch cultures grown with 10 mM CBA or 10 mM fructose, respectively. B) Starving cultures taken from the chemostat with 10 mM CBA and a dilution rate of 0.05 h\(^{-1}\). Values in brackets correspond to the measured proportion of the free circular form in total clc specific DNA. Lane 1, chemostat culture in steady state (4 %); lane 2 and 3, same culture after 16 h (8%) and 64 h (15%) starvation, respectively; lane 4, same culture as lane 3, but DNA isolated according to the plasmid DNA isolation protocol. Autoradiograms were scanned, processed in Adobe Photoshop 6.0, saved and imported into Adobe Illustrator 8.0.1 as 8 bit TIFF files. For convenience of presentation, in panel B the middle sections of the two original images were left out in order to fit both chromosomal bands on the image.

**Effects of cell density and general stress conditions.** GFP fluorescence values of cells growing under steady-state conditions with 0.1 mM CBA and a dilution rate of 0.05h\(^{-1}\) were slightly higher than those grown with 10 mM CBA (Table 4), despite the 80-fold cell-density difference (OD\(_{600}\) of 0.01 and 0.83, respectively). When cells were taken out from chemostats and incubated for 60 h in batch with no carbon added (C-starvation) 95% percentile fluorescence values of both cultures became equal. When the culture originally fed with 10 mM CBA was concentrated tenfold and incubated under conditions of carbon starvation as before, also no significant changes in GFP fluorescence values were observed (not shown). This suggested that expression of the integrase in both growing and starving cells was not a cell density dependent effect.

None of the following stress factors, which are known to cause induction of prophages: UV irradiation, heat shock, osmotic stress in the presence of sodium chloride, and treatment
with ethanol caused a major increase in GFP fluorescence (Table 3) or accumulation of the excised clc form (not shown).

**Transfer of the clc element.** As a final determinant for integrase activity, we analyzed transfer frequencies of the clc element in bacterial filter mating experiments. In matings between *Pseudomonas* sp. strain B13-jim1 as a donor and *P. putida* UWC1 as a recipient the occurrence of rifampin resistant colonies growing on CBA (indicative for recipients that had acquired the clc element) was 2.0, 5.6 and 2.8 times higher after 24, 48 and 72 h of mating time incubation, respectively, when 1 mM CBA was added to the medium as opposed to fructose (Table 5).

When the net increase of the number of transconjugant colonies was calculated for discrete incubation time intervals, the transfer rate of the clc element seemed to have steadily increased with longer incubation periods. Only slight cell growth occurred during mating itself, since the number of donor and recipient cells between individual filters and different time points fluctuated within a factor of three. From these results we conclude that the observed increase in number of transconjugants resulted from new gene transfer events and not from selective propagation of previously produced transconjugants on the supplemented CBA in the plates. Matings performed with donor cells taken from exponentially growing cultures on MM agar without any C-source resulted in at least 100-fold lower frequencies of transconjugant appearance. After 4 h mating $2.3 \times 10^{-6}$ and $1.0 \times 10^{-6}$ transconjugants arose per donor cell grown before mating on CBA and fructose, respectively. Mating experiments were performed twice with similar outcomes.

**TABLE 5.** Transfer of clc element from *Pseudomonas* sp. strain B13-jim1 to *P. putida* UWC1 on membrane filters in the presence of 3-chlorobenzoic acid or fructose

<table>
<thead>
<tr>
<th>Mating time, h</th>
<th>Transfer frequency</th>
<th>Incubation period, h</th>
<th>Transfer rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CBA&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Fructose&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CBA&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>24</td>
<td>$1.27 \times 10^{-3}$ (0.21 x $10^{-3}$)</td>
<td>$0.63 \times 10^{-3}$ (0.34 x $10^{-3}$)</td>
<td>0-24</td>
</tr>
<tr>
<td>48</td>
<td>$50.8 \times 10^{-3}$ (11 x $10^{-3}$)</td>
<td>$9.05 \times 10^{-3}$ (3.0 x $10^{-3}$)</td>
<td>24-48</td>
</tr>
<tr>
<td>72</td>
<td>$249 \times 10^{-3}$ (84 x $10^{-3}$)</td>
<td>$89.8 \times 10^{-3}$ (42 x $10^{-3}$)</td>
<td>48-72</td>
</tr>
</tbody>
</table>

<sup>a</sup>) Transfer frequency (standard deviation within brackets) was calculated as the number of transconjugants per donor cell counted after 24, 48 and 72 h mating incubation time, respectively.

<sup>b</sup>) Both CBA and fructose were present at 1 mM concentration in the filter support medium.

<sup>c</sup>) Transfer rate was calculated as the net increase in the number of transconjugants per donor cell per hour during the corresponding incubation period.
DISCUSSION

The clc element of *Pseudomonas* sp. strain B13 is a representative of a growing class of genomic islands (14, 37). Although genomic islands have mostly been associated with pathogenicity determinants (19), the clc element is devoid of such factors, but its integrase has definitive evolutionary and mechanistic relationships with integrases of pathogenicity islands (37). Contrary to most pathogenicity islands, however, the clc element not only excises, but mediates its own conjugal transfer and reintegration into a new recipient, making it a valuable model system to study the behavior of genomic islands. One of the central features among genomic islands certainly is the integrase, which mediates site-specific integration and excision. From the work presented here we can conclude that expression of the integrase in its integrated (*attR*) situation is tightly controlled and subject to regulation by environmental and growth conditions, with starvation/stationary growth phase being the main trigger. Quite surprisingly, one of the conditions causing induction of the integrase was cultivation on CBA, which is the compound specifically metabolized by enzymes encoded on the clc element. This suggests a signalling or regulatory link between a specific growth substrate (i.e., CBA) and gene transfer rates of the DNA region responsible for this metabolic trait (the clc element), which, in our opinion, points to directed evolution.

In general, transcription from the integrase promoter in the integrated form was tightly repressed resulting in very low levels of GFP production in almost all cells under all growth conditions tested. Whenever induction was observed, it always occurred in only a small subpopulation of cells (at most 13%). This is in contrast to what is mostly described for inducible gene expression, in which case induction will lead to a heterogenous but log normal type of expression patterns in a bacterial population (21). Rather, this behavior is reminiscent of stochastic gene expression observed in some prophages (1). Stochastic gene expression is thought to be the result of a balance between a repressor and activator acting on the same operator/promoter site (1). The conditions causing subtle but most pronounced stimulation of the integrase promoter occurred in stationary phase growth and under starvation conditions (which cannot be easily distinguished), with further enhancement when the cells had been growing on CBA. This we concluded from batch growth and continuous cultivation experiments of the GFP reporter strain B13-jim1 with CBA or fructose as carbon substrate.

Levels of GFP fluorescence of single cells seemed to be representative for induction of the integrase gene itself and not an artifact from placing a separate $P_{\text{ura}}\text{gfp}$ fusion elsewhere in
the chromosome of strain B13. This was inferred from two independent other types of assays for integrase activity: i) the appearance of the excised circular form of the clc genomic island by quantitative hybridization and ii) transfer rates of the clc element in bacterial conjugative matings. In exponentially growing cells the amount of the excised form of the clc element was hardly detectable by Southern hybridization (compared to the chromosomal copies), whereas during the stationary phase the excised form emerged and accumulated. Both in stationary phase cultures and under carbon starvation the abundance of the free circular form was twice as high in cells grown on CBA compared to fructose. Conjugation transfer frequencies of the clc genomic island between B13-jim1 as donor and P. putida UWC1 also increased during matings for up to 72 h, suggesting that stationary phase conditions and/or carbon starvation were a stimulus for the transfer process. Furthermore, the transfer frequencies were up to five times higher in the presence of CBA as opposed to fructose, suggesting that the enhancement of CBA was also visible in the transfer process.

We can currently only hypothesize as to how starvation and CBA exert their effects on the integrase promoter. Most likely, CBA itself does not act as signalling compound for induction, since spiking of CBA to starving cells of strain B13-jim1 taken from the CBA-fed chemostat did not result in production of extra GFP. Autoinducer signalling of the acylhomoserine lactone type did not seem to be involved either, since at different cell densities no effects on GFP expression from the integrase promoter were observed. None of the treatments that typically trigger the SOS/RecA response affected GFP expression or appearance of the circular form, suggesting that a RecA/LexA type degradation is not directly involved in regulating integrase expression of the clc element. Our present hypothesis is that a metabolite of CBA is at some point acting as an effector on modulating the interactions of transcription factors with the intB13 promoter, or on the synthesis of those by interfering with another ‘master’ regulator acting. In fact, additional evidence has been obtained from mutation analysis for the existence of both an activator and repressor acting on the integrase promoter (Chapter 3). Since other examples of subtle triggering, like the inducible effect of some antibiotics on integrase expression of prophages (39) or of the Enterococcus conjugative transposon Tn916 (22, 33) have been described, it seems quite possible that such a direct effector-regulator interaction would be present in the clc element.

What is the evolutionary significance of the clc genomic island? It is the first example of a class of genomic islands now viewed as ‘ecological’ or catabolic genomic islands (14), which may be more widely distributed in microbial communities and responsible for transferring ecological-type functions. In at least two other microorganisms we have found
good evidence for the presence of genomic islands highly similar to the clc element. In the *Xylella fastidiosa* genome sequence (34) we could identify by sequence homologies a region of 71 kb flanked on one end by a glycine tRNA gene and an integrase and on the other end the 18-bp repeat typical for the clc element (37). From partial sequence data of the clc element’s left end, it also became apparent that the clc element and the Xylella element are highly related in this region (Chapter 3), although the Xylella element does not show any genes for chlorobenzoate degradation (34). In *Ralstonia* sp. strain JS705, an environmental isolate from chlorobenzene contaminated groundwater (38), an almost identical clc element was found, which, however, had acquired an extra insertion of about 12 kb (25). These examples show that the clc element (and related types of elements) are capable of being distributed among different bacteria, also in the natural environment, a capacity which is perhaps enhanced by triggering effects of environmental chemicals.
REFERENCES


CHAPTER 5

Concluding remarks

The work presented in this dissertation was focused on different molecular aspects of evolution of catabolic pathways in the genus *Pseudomonas*. In particular, our objectives were to understand the often rapid formation and distribution of metabolic pathways for aromatic compounds, and the possible role that environmental conditions may have on the rates of the distribution. For this purpose, we used two systems: i) the genes for toluene and xylene degradation, mostly located on plasmid DNAs and ii) the genes for 3-chlorobenzoate degradation of *Pseudomonas* sp. strain B13, located on a genomic island.

**Origin and evolution of TOL plasmids.** This part of the thesis clearly demonstrated that the genes for toluene and xylene degradation are distributed as ‘blocks’ or ‘cassettes’ on different plasmids in bacterial isolates of a rather restricted geographical region. Twelve TOL plasmids from Belarus were characterized and found to contain *xyl* genes highly identical to one another and to those of the archetype TOL plasmid pWW0. The *xyl* genes were carried by rather different unrelated plasmid 'vehicles' (Chapter 2). This indicates that natural interreplicon exchange of the *xyl* operons had occurred, in other words, that the complete *xyl* metabolic pathway had been recruited by different plasmid replicons allowing their host cells to selectively proliferate in an environment polluted with toluene and xylene. Our comparative study therefore suggested that TOL plasmids can be regarded as being made of two basic modular structures: (i) the *xyl* catabolic genes and (ii) the plasmid ‘vehicle’.

On a more detailed level of DNA sequence, the *xyl* gene module appeared to be steadily evolving due to DNA recombinations and genetic drift (Chapter 2). It seems amazing that even such pathways as the *meta* cleavage pathway for toluene and xylenes metabolism, which are considered evolutionary ‘old’ (16) are still going through the process of gene reshuffling and 'trying out' different combinations of catabolic genes. These recombinations involved sometimes DNA segments with as low as 70% nucleotide identity and resulted in mosaic catabolic operons, like the upper pathway operons on pWW0 and pSVS plasmids of type C (Chapter 2). Therefore, one could even assume the existence of a pool of evolutionary diverged *xyl* alleles recombining with each other and with other catabolic gene clusters carried on and horizontally distributed by plasmids. Perhaps such a shared horizontal pool of multiple *xyl* alleles enables the formation and selection of those combinations of *xyl* genes
which are most suitable for local conditions, such as the range and availability of particular aromatic growth substrates, temperature, or pH. Probably, transfer of xyl genes and further DNA recombinations can also result in the assembly of new metabolic pathways like, for example, for nitrotoluene (5) or dichlorobenzoate utilization (12) in a single microorganism. Interestingly, at some evolutionary checkpoint, one particularly efficient configuration of the xyl genes evolved, i.e., the contemporary upper and meta-pathway operons plus their regulators, which was subsequently picked up by conjugative and mobilizable plasmids and/or transposons, and distributed among bacteria all over the globe. A similar process appears to have occurred in the dissemination of antibiotic resistance genes by different mobile genetic elements throughout microbial communities. Some xyl gene – plasmid backbone combinations, like, for example, those found on the plasmid pWW53, appeared to be more successful under environmental conditions than others, as judged from their wide geographical distribution and the prevalence of pWW53-type plasmids in different soil samples (Chapter 2).

Analysis of the non-catabolic part of pWW53-type plasmids documented extensive DNA rearrangements, with up to 100 kb or more than half of the total plasmid content, being acquired or lost (Chapter 2). At this point we can only hypothesize about the nature of the variable DNA (except for the mercury resistance genes on pSVS16) and its fate. It is clear, though, that TOL plasmids participate in the processes of DNA acquisition, loss, exchange and horizontal transfer fuelling adaptive evolution of their host. Consequently, at the moment of their isolation TOL plasmids may contain larger regions of dispensable DNA which has only recently been obtained or which has been retained simply due to its physical linkage to selectable markers.

**Evolutionary relationships of the clc genomic island.** The theme of the recruitment of the adaptation genes by mobile genetic elements was further elaborated in this thesis by studying the characteristics of another type of mobile DNA element, the clc genomic island of *Pseudomonas* sp. strain B13 (Chapter 1 and 3). Until a few years ago, the clc genomic island seemed to be a rather sole representative of a mobile DNA form, which was specialized in integrating itself into the bacterial chromosome, but still capable of producing a circular form, from which conjugative transfer to a new recipient could occur. However, recent complete genome sequences produced evidence for many more genomic islands, all of which could be characterized by a conserved integrase gene, attachment site, regulatory region, and putative transfer region - together comprising basic structure of the now so-called integrative and conjugative elements (ICE, (2)). Interestingly, DNA sequence comparisons between several
such genomic islands again indicated some sort of modular structure. About one-third to two-thirds of the size of the genomic islands appeared to contain element-specific functions, like for instance, carbohydrate metabolism in *X. fastidiosa* (13), heavy metal resistance and putative pathogenicity factors on PAGI-2 and PAGI-3 in *P. aeruginosa* (6), chlorobenzoate and chlorocatechol degradation genes on the *clc* island and chlorobenzene oxidation genes on a sibling of the *clc* island in *Ralstonia* sp. isolate JS705 (7, 15). This suggests that genomic islands (and also the *clc* island) keep a relatively strongly constrained region for mechanistic functioning (i.e., the integrase and regulatory genes) and a more flexible region in which new other genes or gene fragments can be acquired or deleted.

This concept of modular structure was extended for all kinds of mobile elements (14), e.g. transposons, plasmids, bacteriophages and ICEs. Some of these functional modules can exchange between ‘classes’ of mobile elements. The *clc* island seems no exception here: it contains a ‘bacteriophage-type’ integration module (i.e. integrase and attachment site), ‘plasmid-type’ maintenance/regulation module (e.g. *parA* and *parB* partitioning-like genes) and several selection modules (i.e. catabolic genes, found separately in other bacterial isolates). We can envision that the *clc* element originated via fusion of a self-transmissible plasmid replicon with a bacteriophage. The *clc* element still seems to have kept its ability to replicate, which we conclude from the fact that after conjugative transfer to recipients carrying two integration sites always two copies of the island occur (11).

**Survival strategy of the *clc* island and regulation of the horizontal gene transfer by environment.** The most surprising findings of this thesis, but most difficult to interpret concern the regulation of transfer of the *clc* element. In growing cultures of *Pseudomonas* sp. strain B13 the island was present exclusively in its integrated form, which correlated with extremely low transfer frequencies (Chapter 4). In stationary growth phase and under conditions of carbon starvation the *clc* island became activated, excised from its chromosomal location and was transferred to new recipients, albeit still at most in about 10% of the cells. The rate of transfer under stationary phase conditions increased 10,000 fold compared to exponentially growing cultures (Chapter 4). We currently interprete this as if the island ‘escapes’ to another bacterial host via cell-to-cell contact under adverse conditions. This strategy of the *clc* island resembles that of temperate bacteriophages which enter the lytic cycle upon unfavorable environmental circumstances apparently in order to rescue phage progeny (e.g. (9, 10). The *clc* element is clearly different in behavior from bacteriophages in that it is not susceptible to activation by UV irradiation, heat shock or chemical stress. Strangely enough though, induction of the *clc* island appeared to be stimulated at least
twofold after growing cells in the presence of 3-chlorobenzoic acid compared to other growth substrates. We inferred this from analysis of induction from a central promoter on the \textit{clc} element, the integrase promoter, from measuring the appearance of the excised form, and from measuring transfer rates of the \textit{clc} element. This suggests that there might be a link between 3-chlorobenzoate metabolism and the rates of transfer of the \textit{clc} element. The nature of this link is currently not clear. It might be that 3-chlorobenzoate or a metabolite causes some indirect effect on a transcription factor like RpoS, which is specifically present in stationary phase and might influence expression of the integrase promoter. On the other hand, it is still possible that 3-chlorobenzoate or a metabolite influence expression of the integrase promoter more directly by interacting with a specific transcription regulator of integrase expression. As described in Chapter 3 and 4, the stimulating effect by 3-chlorobenzoate is unique in its kind, although other chemical effectors are known to influence conjugation transfer rates. Some of the best examples of these effectors are formed by the \textit{Agrobacterium} tumor-inducing plasmids which are stimulated in the presence of plasmid-specific opines (4, 8). The other case is the stimulation of transfer of conjugative transposons by tetracycline. In contrast to the stimulatory effect of 3-chlorobenzoate, which –at best- is around a factor of 2, tetracycline and opines can stimulate conjugative transfer by a factor of 10,000.

As far as we could study in this dissertation, the integrase is a key function determining the behavior of the \textit{clc} island. We could establish that the \textit{intB13} gene is expressed from two alternative promoters depending on whether the island is in its circular excised form or integrated. In the circular form, expression occurs constitutively from a relatively strong promoter, but in the integrated form transcription of the integrase is under stationary phase/starvation and C-source control. Regulation of integrase expression from the integrated form seems to involve the interplay of activation and repression mechanisms (Chapter 3 and 4), some of which could be identified as being encoded on the \textit{clc} element itself. This suggested that regulation of integrase expression is cell-independent to some level and coordinated by the element itself. The interplay of activating and repression mechanisms on the same promoter could result in this strange phenomenon of subpopulations with induced integrase promoters. When assuming that activator and repressor proteins are not distributed equally among all cells in a population, each individual cell may carry a different ratio of activator to repressor. If successful transcription of the integrase gene depends on the ratio of the activator to repressor level, one can imagine that only in a subpopulation of cells (namely those in which the ratio accidentally is favourable enough) integrase expression occurs. This effect was modeled by others and termed stochastic gene expression (1, 3).
Future research will have to resolve further structural and functional aspects of the regulation of excision and conjugative transfer of the clc island. It seems particularly interesting to find out if 3-chlorobenzoate or a metabolite are really interacting with one of the transcription activators of the integrase gene and thus directly stimulating transfer of the genes for their own degradation (although it might be hard for us to see the ‘purpose’ of that). In addition, several features important for the life style of the clc island are worthy to be studied further. How is it achieved that donor cells always seem to keep two copies of the clc element, even if some donate the element to new recipients? Is perhaps transfer occurring via single stranded DNA from a rolling circle, whereas the other strand is again replicated in the donor cell and reintegrated in the chromosome. How is the conjugative process mediated? Preliminary sequence analysis of the complete clc island showed very little homology to known plasmid conjugative systems, which might indicate a novel sort of conjugation. Finally, an investigation on the occurrence of clc type genomic islands among various bacteria will give a more comprehensive picture on their ecological significance and their ability to acquire and distribute catabolic and other functions within microbial communities.

REFERENCES


CURRICULUM VITAE

Name
Vladimir Sentchilo

Date of birth
June, 25, 1972

Place of birth
City of Minsk, Belarus

1979-1989
Primary and secondary school No.111, Minsk, Belarus.

1989-1994

1994-1997
PhD studies at the Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk. Research topic: Polymorphism of plasmids encoding catabolism of toluene in Pseudomonas. Minsk, Belarus.

Research fellow at Swiss Federal Institute for Environmental Science and Technology (EAWAG), Dübendorf, Switzerland.

PhD studies at Swiss Federal Institute for Environmental Science and Technology (EAWAG), Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, under supervision of Dr. J. R. van der Meer and Prof. Dr. A. J. B. Zehnder

26/6/2003
Date of the oral examination
LIST OF PUBLICATIONS


