Indications for Acquisition of Reductive Dehalogenase Genes through Horizontal Gene Transfer by *Dehalococcoides ethenogenes* Strain 195

Christophe Regeard,1† Julien Maillard,1 Christine Dufraigne,2 Patrick Deschavanne,2‡ and Christof Holliger1*

Swiss Federal Institute of Technology Lausanne (EPFL), ENAC-Laboratory for Environmental Biotechnology, 1015 Lausanne, Switzerland,1 and INSERM U494, 91 bd de l’Hôpital, 75634 Paris Cedex 13, France2

Received 6 May 2004/Accepted 17 December 2004

The genome of *Dehalococcoides ethenogenes* strain 195, an anaerobic dehalorespiring bacterium, contains 18 copies of putative reductive dehalogenase genes, including the well-characterized *tceA* gene, whose gene product functions as the key enzyme in the environmentally important dehalorespiration process. The genome of *D. ethenogenes* was analyzed using a bioinformatic tool based on the frequency of oligonucleotides. The results in the form of a genomic signature revealed several local disruptions of the host signature along the genome sequence. These fractures represent DNA segments of potentially foreign origin, so-called atypical regions, which may have been acquired by an ancestor through horizontal gene transfer. Most interestingly, 15 of the 18 reductive dehalogenase genes, including the *tceA* gene, were found to be located in these regions, strongly indicating the foreign nature of the dehalorespiration activity. The GC content and the presence of recombine-nase genes within some of these regions corroborate this hypothesis. A hierarchical classification of the atypical regions containing the reductive dehalogenase genes indicated that these regions were probably acquired by several gene transfer events.

The increasing knowledge of gene and genome sequences provides strong evidence that horizontal gene transfer (HGT) plays a major role in the evolution of the prokaryotic genomes. Primarily evoked due to the emergence of multidrug resistance patterns (4, 7), it is now well accepted that acquisition and integration of new DNA in prokaryote chromosomes are dynamic processes allowing rapid adaptation to new ecological niches (11). Sequenced bacterial genomes have revealed that a significant part of the genomes originates from HGT (43). For *Escherichia coli*, it has been estimated that 24% of the entire genome has been acquired by HGT (30). In contrast to the core gene pool, the horizontal gene pool (54), also called the flexible gene pool (15), is composed mainly of operational genes rather than informative genes (47). The exchange of the horizontal gene pool between bacteria sharing a common ecological niche is mediated by transmissible mobile genetic elements (MGEs). Numerous phages, plasmids, transposons, integrons, and genomic islands participating in DNA transfer have been described and can be considered as a mosaic continuum (44).

In the last few decades, the environment has been polluted by numerous synthetic chemicals that are foreign, and potentially dangerous, to life either by their structure or by their unnaturally high concentrations. Despite their xenobiotic character, many of these compounds have been found to be biodegradable. However, in some environments, it can take weeks, months, and even years for significant biodegradation to be observed. This rather short time period on an evolutionary scale has led to speculations on the rapid evolution of catabolic pathways. Current knowledge on the role of MGEs in bacterial evolution has been recently reviewed (55, 57, 58). Different types of experiments have revealed the existence of the metabolic horizontal gene pool. Transmissible MGEs containing degradation genes have been found for toluene-xylene (50), naphthalene (18), and chlorobenzene (38) degradation, to name a few. Direct evidence of natural spreading of MGEs in the ecosystem has been obtained when bacteria with MGEs have been used for bioremediation of polluted sites by bioaugmentation (45, 56). Indirect evidence for HGT has often been illustrated by incongruities in phylogenetic trees, as demonstrated for dissimilatory sulfite reductase genes of sulfate-reducing bacteria (26) and ferredoxins, nitroreductases, NADH oxidases, and alcohol dehydrogenases of anaerobic prokaryotes and microaerophilic protists (41).

When orthologous sequence comparisons have not been possible, other approaches based on atypical features of a sequence in its genomic context, such as nucleotide composition (29), relative frequencies of dinucleotides (25), codon usage biases (37), and Markov chain pattern analyses (16), have been developed. These approaches revealed indications of HGT of a glycosyl hydrolase between bacteria and fungi in the rumen (13) and of UDP-glucose dehydrogenase between *Streptococcus* sp. and *Escherichia coli* (39). More recently, a new tool allowing an easy screening of HGT in genomes has been developed based on signatures of sequences (9). The genomic signature concept shows that oligonucleotide usage varies less along a genome than between genomes and is therefore species specific (8, 24, 48). Using this concept, local variations in the genomic signature of a particular genome re-
vealed regions that contain an atypical signature and therefore
could originate from HGT (C. Dufraigne, B. Fertil, S. Lespinats, A. Giron, and P. Deschavanne, unpublished data). The
aim of this work was to study the genome of *Dehalococcoides ethenogenes* strain 195 (The Institute for Genomic Research [TIGR] database) by analyzing variation of the genomic sig-
nature along the genome and using other bioinformatic tools in
order to obtain indication of HGT events. *D. ethenogenes* strain 195 is a bacterium that depends strictly on chloroethenes (tet-
rachloroethene [PCE], trichloroethene [TCE], and dichlo-
roethene) and 1,2-dichloroethane as an electron acceptor and
uses hydrogen only as an electron donor (12, 36). *D. etheno-
genes* strain 195 is the first bacterium described in pure culture
that is able to completely dechlorinate PCE to ethene in a
process called dehalorespiration for the first dechlorination
steps and a cometabolic process for the last step (19). At least
two reductive dehalogenases are involved in the dechlorination
to vinyl chloride, one dechlorinating PCE to TCE and the
other dechlorinating TCE to vinyl chloride (32). For the latter,
the gene *tceA* has been cloned and sequenced (31), showing
the presence of a Tat signal peptide in the N-terminal region
and motifs for the binding of two iron-sulfur clusters in the
C-terminal region. In the genome of *D. ethenogenes* strain 195,
16 to 17 putative reductive dehalogenase genes (*rdhA*) have
been identified (references 51 and 60, respectively). The
present study, which examines the variation of the genomic
signature in the genome of *D. ethenogenes* strain 195, indicates
the foreign nature of nearly all of these putative reductive
dehalogenases. Open reading frames (ORFs) at the borders of
the detected foreign genome regions (referred to as atypical
regions) showed good similarity with recombinase genes (a
generic term used here that includes transposases, insertion
elements, integrases, and recombinases involved in DNA mo-
bility [42]). The comparison of the genomic signature of
the different atypical regions indicates that different HGT events
have occurred which led to this large pool of putative *rdhA*
genes.

**MATERIALS AND METHODS**

**Genome DNA sequence data.** Preliminary sequence data of the genome of *D. ethenogenes* strain 195 were obtained under license from TIGR through their website (http://www.tigr.org). The genome version used in this study was the one from September 2002 (contig 6871). As pointed out by TIGR, it is not excluded that several sequencing errors exist in this version and will be corrected for the final publication of this genome. However, these possible errors have little consequence for the results presented in this study.

**Genomic signature determination.** Genomic signatures were obtained as de-
scribed previously (9). Briefly, the fast-algorithm Chaos Game Representation (23) (a program developed with C++ and Matlab) allows treatment of 1-Mb sequences in a few seconds on a laptop computer and calculation of the whole set of frequencies of short oligonucleotides found in a sequence. Signatures are then visualized as square images where the color of each pixel represents the fre-
cency of a given oligonucleotide (later referred to as “word”).

**Detection of atypical regions.** The detection of atypical regions, possibly origi-
nating from HGT, is based on the determination of differences between local
signatures and the signature of the genome as a whole (Dufragne et al., unpublished). The genome of *D. ethenogenes* was analyzed for local signatures using the following optimal parameters: sampling of the genome by a sliding window of 5 kb (with steps of 2.5 or 0.5 kb) in which the frequency of four-letter words is
determined. The strand asymmetry bias was ruled out using double-strand DNA
for the analysis. The distinction between the intrinsic or host signature of a
bacterium, which contains all the vertically inherited genes, and the signature
of atypical regions required a precise analysis of the signatures obtained along the
genome. For that purpose, the preponderant signature which may represent the
host signature was determined with the K-means clustering algorithm. The host
signature cluster is the one that has the largest sample size and is independent
of the number of cluster seeds. The comparison between the signature of a window
and the host signature was obtained by a Euclidian metric algorithm accounting
for differences in word usage (Dufragne et al., unpublished).

**Analysis of atypical regions.** The gene content of detected atypical regions was
studied with the following strategy: (i) the DNA corresponding to the atypical
region was separated from the rest of the genome; (ii) the program ORF finder
(http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to obtain a map with all
possible ORFs present in the six existing frames; and (iii) ORFs with a length of
≥500 bp were searched for sequence homology with either Cognitor (http://www
.ncbi.nlm.nih.gov/gorf/orfg.cgi), which allows comparison with COG (cluster
of orthologous groups of proteins) or, when a protein family could not be
determined, BLASTp (2). The GC content of each studied atypical region was
determined by the GeeCee program (http://bioweb.pasteur.fr/seqanal/interfaces
/geecee.html).

**Comparison of genomic signatures.** The signatures consisting of the frequency
matrices of all four-letter words were compared with each other using a hierar-
chical clustering (JMP software; SAS Institute Inc.). Briefly, this clustering
groups the closest signatures together, calculates their mean values, and iterates
the process for all signatures.

**RESULTS**

**Detection of atypical regions in the genome of *D. ethenogenes*.** The version of the *D. ethenogenes* strain 195 genome used here consisted of a major contig (numbered 6871) of a
size of 1,470,272 nucleotides. The GC content of this small genome is 48.85%. In order to investigate the possibility that
this bacterium has acquired the putative reductive dehalogene-
nase genes by horizontal gene transfer, the local genomic sig-
nature along the genome was calculated as described in Ma-
terials and Methods. By reading the genome with small sliding
windows (5 kb) from the beginning to the end of the contig,
local differences in genomic signature were detected (Fig. 1A).
By visual analysis, it was possible to recognize that at different
places in the genome, the apparent homogeneity of the picture
was disrupted. For a more accurate analysis of the positions of
these fractures, a clustering in four classes was performed (Fig.
1B). The most populated class, referred to as class I, accounts
for 78% of the genome and represents the intrinsic or host
signature of *D. ethenogenes* strain 195. The three other classes
that account for 22% of the genome are good candidates for
DNA acquired by HGT. Classes II, III, and IV account for
17%, 4.9%, and 0.1% of the genome, respectively. It is impor-
tant that the different classes have no real genetic signification,
they only represent parts of the genomic DNA that fall in the
same cluster of signatures.

**Analysis of atypical regions.** The gene content analysis of the atypical regions of class III indicated that a putative tem-
perate bacteriophage is present in the genome of *D. etheno-
genes* strain 195. The large atypical region located around 550
to 600 kb (Fig. 1B) contains only genes that are related to
phage DNA. The GC content of this putative phage is identical
to the GC content of its host. This shows the usefulness
of signature analysis for the detection of foreign DNA, since a
simple GC content analysis would not have revealed the puta-
tive prophage. The genes present in the second half of the
atypical regions A and H (designated A2 and H2, respectively)
that cluster with class III also display significant homology with
the phage protein family. As previously described (Dufragne
et al, unpublished), rRNA genes have their own specific sig-
nature because of their very low mutation rate and therefore
are located in atypical regions. In the *D. ethenogenes* strain 195
genome, the gene encoding the 16S rRNA (GenBank accession number AF004928) was localized in an atypical region around position 642700. The gene encoding the 23S rRNA was localized around position 52000 in a small atypical region with a signature different from the host signature. The class IV atypical region, which accounts for only 0.1% of the genome and showed the largest distance from the host signature, corresponds to an atypical region containing ORFs without clear function but with high alanine content. This region can be qualified as a low-complexity region probably due to a high number of short repetitions.

Identification and analysis of putative reductive dehalogenase genes. A BLAST sequence comparison (TBlastn) of the genome of *Dehalococcoides ethenogenes* strain 195 with the protein sequence of the characterized reductive dehalogenase TceA as a query retrieved 17 genes with significant homology (26 to 32%). The characteristics of the putative reductive dehalogenase genes (referred to as rdhA-De1 to rdhA-De17) are summarized in Table 1. The N-terminal sequences are given to allow a clear identification of each gene after annotation of the genome by TIGR. A BLAST2 sequence analysis revealed that homology between all 18 rdhA gene products ranged between 24 and 54% identity (data not shown). This strongly suggests that duplication phenomena and evolution by point mutation can be excluded as explanations for the high number of putative reductive dehalogenase genes. The 18 rdhA genes possess the typical features of reductive dehalogenases, e.g., a Tat signal peptide in the N-terminal part and two iron-sulfur cluster binding mo-
of the atypical regions detected by the signature analysis. Fifteen genes, one of which is the dehalogenase gene, encoding a small hydrophobic protein with two or three transmembrane helices (40, 52). A membrane anchor function has been proposed for this protein, and co-transcription of the two genes has been shown by reverse transcription-PCR (40, 52). A rdhB gene was found downstream of the rdhA gene for all putative reductive dehalogenase genes of D. ethenogenes strain 195. The orientation of the rdhA and rdhB genes seems to be conserved among the two main classes of reductive dehalogenases. While they appear as rdhAB cluster in chloroethene reductive dehalogenases, the opposite orientation (rdhBA) is found in chlorophenol reductive dehalogenase clusters (34, 60).

**Localization of the putative reductive dehalogenases.** The location of the rdhA genes was matched with the location of the atypical regions detected by the signature analysis. Fifteen of the 18 rdhA genes were found in atypical regions. Five genes, one of which is tceA, are located alone in atypical regions A, B, D, E, and G, respectively (Fig. 2). The genes rdhA-De4 to rdhA-De7, rdhA-De10 to rdhA-De13, and rdhA-De16 and rdhA-De17 are grouped and located in the atypical regions B and C, respectively. Interestingly, rdhA-De17, lacking the N-terminal signal peptide, is located in H2. Genes rdhA-De2 and rdhA-De3 and gene rdhA-De15 are found in regions exhibiting the host signature, between atypical regions B and C and between regions G and H, respectively. Atypical regions A to H have a GC content clearly under 50%, a value that is different from that of the host (49.54%) (Fig. 2).

**Genetic characterization of the atypical regions.** A detailed analysis of the upstream and downstream regions of the cprBA gene cluster of Desulfotobacterium dehalogenans, an o-chloro-phenol-dehalorespiring bacterium, revealed several genes encoding proteins that are possibly involved in either the regulation of transcription of cprBA (CprC and CprK) or the correct folding, processing, and assembly of the reductive dehalogenase (CprD, CprE, and CprT) (52). By sequence comparison, the genome of D. ethenogenes strain 195 was screened for the presence of homologous genes. Four genes putatively encoding CprC-like proteins were identified. Two of these putative rdhC genes have no apparent link with rdhAB genes, whereas the other two are located in atypical region F between rdhA-De11 and rdhA-De12 and in atypical region G directly upstream of rdhA-De14, respectively. These putative rdhC genes have the same transcription direction as the rdhAB genes. Three genes that can potentially encode proteins homologous to CprE, CprD, and CprT, respectively, have been found, but all three are not present in an atypical region or in the vicinity of an rdhAB cluster. Finally, one gene possibly encoding a CprK-like protein was localized in the atypical region F directly downstream of rdhA-De13 (data not shown).

In order to obtain additional evidence that the atypical regions A to H are foreign DNA acquired by HGT, they were analyzed for the presence of genes encoding putative recombinases typically present in transmissible MGEs. With the exception of the atypical regions C, D, and G, all other atypical regions contain recombinase-like genes (indicated as a grey arrow in Fig. 2). In atypical region H, consisting of two parts with class II and class III signatures, the recombinase-like genes are located in H2 belonging to class III (resembling the phage signature).

Interestingly, all rdh genes located in the first half of contig 6871 (atypical regions A to E) have the same transcription direction. The same pattern was observed for the second half, however, in the opposite direction. This suggests that the origin of replication is located somewhere between atypical regions A and H. Therefore, the low-complexity region associated with

<table>
<thead>
<tr>
<th>Putative reductive dehalogenase gene</th>
<th>N-terminal amino acid sequence</th>
<th>Length of deduced protein (aa)</th>
<th>Position in genome</th>
<th>Location in atypical region</th>
</tr>
</thead>
<tbody>
<tr>
<td>tceA-De</td>
<td>MSEKYHSTTVTRR</td>
<td>554</td>
<td>23945</td>
<td>A</td>
</tr>
<tr>
<td>rdhA-De1</td>
<td>MSSFHISTVSRRRD</td>
<td>482</td>
<td>166788</td>
<td>B</td>
</tr>
<tr>
<td>rdhA-De2</td>
<td>MSKFHSMVSRDD</td>
<td>500</td>
<td>178825</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De3</td>
<td>MNKFHSTLSRRRD</td>
<td>492</td>
<td>184414</td>
<td>C</td>
</tr>
<tr>
<td>rdhA-De4</td>
<td>MKKHSTLSRRRD</td>
<td>494</td>
<td>185075</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De5</td>
<td>MSKFHSTVSRDD</td>
<td>469</td>
<td>191688</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De6</td>
<td>MSKLHSTLSRRRD</td>
<td>507</td>
<td>196819</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De7</td>
<td>MNQFHHSTVSRDD</td>
<td>505</td>
<td>199588</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De8</td>
<td>MTEVNRDFDLKA</td>
<td>532</td>
<td>503166</td>
<td>D</td>
</tr>
<tr>
<td>rdhA-De9</td>
<td>MSNFHSTVSRDD</td>
<td>510</td>
<td>767390</td>
<td>E</td>
</tr>
<tr>
<td>rdhA-De10</td>
<td>MLNFHSTLTKRD</td>
<td>495</td>
<td>1266422</td>
<td>F</td>
</tr>
<tr>
<td>rdhA-De11</td>
<td>MKKFHSTLSRRDD</td>
<td>515</td>
<td>1272322</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De12</td>
<td>MSQKHSVTSRDR</td>
<td>505</td>
<td>1276843</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De13</td>
<td>MNKFHIVSRRDD</td>
<td>514</td>
<td>1280982</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De14</td>
<td>MKNFHLSTLSRRDD</td>
<td>490</td>
<td>1344813</td>
<td>G</td>
</tr>
<tr>
<td>rdhA-De15</td>
<td>MDKKNRFDFVK</td>
<td>455</td>
<td>1397826</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De16</td>
<td>MKNFHAFTVSRD</td>
<td>510</td>
<td>1403184</td>
<td>–e</td>
</tr>
<tr>
<td>rdhA-De17</td>
<td>MNDGFQVAQEOY</td>
<td>350</td>
<td>1413295</td>
<td>–e</td>
</tr>
</tbody>
</table>

**a** The position of the first base of each rdhA gene is given in accordance to the genome version used in this study.

**b** The location of the atypical regions in the genome is depicted in Fig. 1.

**c** The genes rdhA-De2 and rdhA-De3 are located between atypical regions B and C; rdhA-De15 is located between regions G and H.
the class IV signature may represent the terminus of replication of *D. ethenogenes* strain 195 genome (Fig. 1B).

**Signature comparison of the atypical regions.** Since the atypical regions were clustered in an arbitrarily defined number of classes, their relationships to each other were further considered. The signatures of the atypical regions, together with those of two short host regions (60 to 70 kb and 1,150 to 1,200 kb) and the integrated prophage, were subjected to a hierarchical clustering. First, the result confirmed the clustering in four classes, with one class (class IV) not included in the clustering (Fig. 3). Both regions A2 and H2 were similar to the phage region, with all of these regions forming the most distant branch. Both host signatures clustered nicely together. The atypical regions were found to be variably related to each other. Whereas regions E and F and also C and G were closely related despite their wide distribution over the genome, the other regions were found to be more distantly related to each other.

**DISCUSSION**

HGT is considered to be a major factor for the rapid adaptation of bacteria to new ecological niches and environmental stresses (11, 27, 43). Several key findings based on the phylogeny of catabolic genes have evidenced the important role of transmissible MGEs in the acquisition of new catabolic properties (55). The intriguing number of putative reductive dehalogenase genes in *Dehalococcoides ethenogenes* strain 195 (51, 60) and the quite recent accumulation of chlorinated xenobiotics in the environment raised the question of the evolution of these genes. Starting with unfinished genome data from TIGR and prior to annotation, the global genome organization of *D. ethenogenes* was analyzed using the local variations of genomic signatures. The genomic signature (Fig. 1A) was obtained by plotting the frequency of all four-letter words of 5-kb sliding windows along the genome. The clustering in an arbitrarily defined number of four classes revealed the presence of at least as many significantly different signatures (Fig. 1B). While the most abundant class represents host DNA (class I [78% of the total genome]), two other classes may be considered as DNA of foreign nature, so-called atypical regions. Whereas the class II signature (17%) is widely distributed over the whole genome, class III (4.9%) is dominated by a large DNA fragment probably consisting of an integrated prophage genome. Class IV DNA may represent the genomic terminus of replication. Interestingly, atypical regions A and H (Fig. 1B) are composed of a mixture of class II and III signatures (A1 and H1 and A2 and H2, respectively). By looking for the exact location of the

---

**FIG. 2.** Locations of *rdh* and recombinase genes in atypical regions A to H. Black arrows represent *rdhAB* genes and illustrate the direction of transcription in the atypical region. Grey arrows represent recombinase genes. Straight lines represent atypical regions belonging to class II, and broken lines represent the atypical region belonging to class III. Numbers represent the positions (kb) of atypical regions in the genome. The GC content of the atypical region is given in parentheses.

**FIG. 3.** Signature comparison of atypical regions containing *rdhAB* genes in the *D. ethenogenes* strain 195 genome. The atypical regions are reported as A to H, and regions A and H are split into two parts (A1/A2 and H1/H2). The signatures of two small host regions and of the detected integrated prophage were included for comparison. Corresponding clustering classes are also indicated.
putative reductive dehalogenase genes, 83% of them, including the well-characterized functional \textit{tceA} gene, were found in atypical regions, 14 in class II and only 1 (\textit{rdhA-De417}, a truncated copy) in class III. Only three copies (\textit{rdhA-De2}, \textit{rdhA-De3}, and \textit{rdhA-De415}) are localized in the host DNA. The presence of various recombinases in most atypical regions and the lower GC content corroborate the hypothesis that these regions may have been acquired by horizontal gene transfer by an ancestor of \textit{D. ethenogenes} strain 195. The recombinase-like genes identified here were shown to have variable homology to ISDatabase sequences (http://www.is.biotoul.fr), ranging from 26% (\textit{rec8}) to 61% (\textit{rec1} and \textit{rec2}) identity. Interestingly, a full insertion sequence was identified directly upstream of \textit{tceA} comprising the \textit{rec1} and \textit{rec2} genes (Fig. 2). This insertion element has strong homology to members of the IS3 family (such as \textit{ISrso14} from \textit{Ralstonia solanacearum} and \textit{IS511} from \textit{Caulobacter crescentus}).

Signature comparisons of the atypical regions (Fig. 3) confirmed the result obtained by the clustering and clearly indicated that regions of class II and class III are different from class I (host signature) and that A2 and H2 are closely related to the phage signature. When speculating about the origin of the atypical regions, at least three distinct events occurred for the integration of regions A1, E, F, and D in one horizontal gene transfer; regions B, H1, C, and G in a second horizontal gene transfer; and the phage cluster in a third horizontal gene transfer. Further rearrangements of DNA within the genome may explain the distribution and signature variations of these regions.

Several bacterial dehalogenases have already been reported to be associated with recombinases, most often insertion sequences. Among them, the best characterized are the haloalcohol dehalogenases of \textit{Dehalococcoides} sp. strain 195. The recombinase-like genes identified here were shown to have variable homology to ISDatabase sequences (http://www.is.biotoul.fr), ranging from 26% (\textit{rec8}) to 61% (\textit{rec1} and \textit{rec2}) identity. Interestingly, a full insertion sequence was identified directly upstream of \textit{tceA} comprising the \textit{rec1} and \textit{rec2} genes (Fig. 2). This insertion element has strong homology to members of the IS3 family (such as \textit{ISrso14} from \textit{Ralstonia solanacearum} and \textit{IS511} from \textit{Caulobacter crescentus}).

Signature comparisons of the atypical regions (Fig. 3) confirmed the result obtained by the clustering and clearly indicated that regions of class II and class III are different from class I (host signature) and that A2 and H2 are closely related to the phage signature. When speculating about the origin of the atypical regions, at least three distinct events occurred for the integration of regions A1, E, F, and D in one horizontal gene transfer; regions B, H1, C, and G in a second horizontal gene transfer; and the phage cluster in a third horizontal gene transfer. Further rearrangements of DNA within the genome may explain the distribution and signature variations of these regions.

Several bacterial dehalogenases have already been reported to be associated with recombinases, most often insertion sequences. Among them, the best characterized are the haloalcohol dehalogenases of \textit{Dehalococcoides} sp. strain 195. The recombinase-like genes identified here were shown to have variable homology to ISDatabase sequences (http://www.is.biotoul.fr), ranging from 26% (\textit{rec8}) to 61% (\textit{rec1} and \textit{rec2}) identity. Interestingly, a full insertion sequence was identified directly upstream of \textit{tceA} comprising the \textit{rec1} and \textit{rec2} genes (Fig. 2). This insertion element has strong homology to members of the IS3 family (such as \textit{ISrso14} from \textit{Ralstonia solanacearum} and \textit{IS511} from \textit{Caulobacter crescentus}).

Signature comparisons of the atypical regions (Fig. 3) confirmed the result obtained by the clustering and clearly indicated that regions of class II and class III are different from class I (host signature) and that A2 and H2 are closely related to the phage signature. When speculating about the origin of the atypical regions, at least three distinct events occurred for the integration of regions A1, E, F, and D in one horizontal gene transfer; regions B, H1, C, and G in a second horizontal gene transfer; and the phage cluster in a third horizontal gene transfer. Further rearrangements of DNA within the genome may explain the distribution and signature variations of these regions.

Several bacterial dehalogenases have already been reported to be associated with recombinases, most often insertion sequences. Among them, the best characterized are the haloalcohol dehalogenases of \textit{Dehalococcoides} sp. strain 195. The recombinase-like genes identified here were shown to have variable homology to ISDatabase sequences (http://www.is.biotoul.fr), ranging from 26% (\textit{rec8}) to 61% (\textit{rec1} and \textit{rec2}) identity. Interestingly, a full insertion sequence was identified directly upstream of \textit{tceA} comprising the \textit{rec1} and \textit{rec2} genes (Fig. 2). This insertion element has strong homology to members of the IS3 family (such as \textit{ISrso14} from \textit{Ralstonia solanacearum} and \textit{IS511} from \textit{Caulobacter crescentus}).

Signature comparisons of the atypical regions (Fig. 3) confirmed the result obtained by the clustering and clearly indicated that regions of class II and class III are different from class I (host signature) and that A2 and H2 are closely related to the phage signature. When speculating about the origin of the atypical regions, at least three distinct events occurred for the integration of regions A1, E, F, and D in one horizontal gene transfer; regions B, H1, C, and G in a second horizontal gene transfer; and the phage cluster in a third horizontal gene transfer. Further rearrangements of DNA within the genome may explain the distribution and signature variations of these regions.

Several bacterial dehalogenases have already been reported to be associated with recombinases, most often insertion sequences. Among them, the best characterized are the haloalcohol dehalogenases of \textit{Dehalococcoides} sp. strain 195. The recombinase-like genes identified here were shown to have variable homology to ISDatabase sequences (http://www.is.biotoul.fr), ranging from 26% (\textit{rec8}) to 61% (\textit{rec1} and \textit{rec2}) identity. Interestingly, a full insertion sequence was identified directly upstream of \textit{tceA} comprising the \textit{rec1} and \textit{rec2} genes (Fig. 2). This insertion element has strong homology to members of the IS3 family (such as \textit{ISrso14} from \textit{Ralstonia solanacearum} and \textit{IS511} from \textit{Caulobacter crescentus}).


