Full Paper

Comparison of the complete genome sequences of four γ-hexachlorocyclohexane-degrading bacterial strains: insights into the evolution of bacteria able to degrade a recalcitrant man-made pesticide

Michiro Tabata, Satoshi Ohhata, Yuki Nikawadori, Kouhei Kishida, Takuya Sato, Toru Kawasumi, Hiromi Kato, Yoshiyuki Ohtsubo, Masataka Tsuda, and Yuji Nagata*

Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Sendai 980-8577, Japan

*To whom correspondence should be addressed. Tel: +81 22-217-5682. Fax: +81 22-217-5682.
E-mail: aynaga@ige.tohoku.ac.jp

Edited by Prof. Takashi Ito

Received 25 April 2016; Accepted 9 July 2016

Abstract

γ-Hexachlorocyclohexane (γ-HCH) is a recalcitrant man-made chlorinated pesticide. Here, the complete genome sequences of four γ-HCH-degrading sphingomonad strains, which are most unlikely to have been derived from one ancestral γ-HCH degrader, were compared. Together with several experimental data, we showed that (i) all the four strains carry almost identical linA to linE genes for the conversion of γ-HCH to maleylacetate (designated “specific” lin genes), (ii) considerably different genes are used for the metabolism of maleylacetate in one of the four strains, and (iii) the linKLMN genes for the putative ABC transporter necessary for γ-HCH utilization exhibit structural divergence, which reflects the phylogenetic relationship of their hosts. Replicon organization and location of the lin genes in the four genomes are significantly different with one another, and that most of the specific lin genes are located on multiple sphingomonad-unique plasmids. Copies of IS6100, the most abundant insertion sequence in the four strains, are often located in close proximity to the specific lin genes. Analysis of the footprints of target duplication upon IS6100 transposition and the experimental detection of IS6100 transposition strongly suggested that the IS6100 transposition has caused dynamic genome rearrangements and the diversification of lin-flanking regions in the four strains.

Key words: xenobiotics, evolution, sphingomonads, genome, mobile genetic elements

1. Introduction

In the early 20th century, rapid development in the chemical industry led to the production and wide use of numerous anthropogenic chemicals. Because they are often recalcitrant in the environment and toxic to humans and ecosystems, they have caused serious environmental problems.1–3 Many bacterial strains capable of degrading man-
made xenobiotic compounds have been isolated and characterized.\textsuperscript{4–9} Such strains are thought to have evolved to degrade xenobiotics within relatively short periods. However, with the exception of a few speculative examples, the evolutionary processes of these bacterial strains remain largely unknown.\textsuperscript{5,10–13} \textsuperscript{c}Hexachlorocyclohexane (\textsuperscript{c}-HCH; also known as \textsuperscript{c}-BHC or lindane) is a completely man-made chlorinated pesticide that has caused serious environmental problems due to its toxicity and long persistence in upland soils.\textsuperscript{7,14,15} Only 60 years after the first release of \textsuperscript{c}-HCH into the environment, a number of bacterial strains that aerobically degrade \textsuperscript{c}-HCH have been isolated from geographically distant locations around the world.\textsuperscript{7} An archetypal \textsuperscript{c}-HCH-degrading strain, \textit{Sphingobium japonicum} UT26, was isolated from an upland experimental field to which \textsuperscript{c}-HCH had been applied once a year for 12 years, and its aerobic \textsuperscript{c}-HCH degradation pathway (Fig. 1) and genome organization have been intensively studied.\textsuperscript{16–19} Recently, many draft genome sequences of other HCH (including not only \textsuperscript{c}-HCH but also other HCH isomers) degraders and their related but non-HCH-degrading strains were determined, and their comparative analyses have been published.\textsuperscript{20,21} These studies provided us some important primary information on the evolution of HCH-degraders with the involvement of plasmids and insertion sequences (ISs). However, the more detailed information (e.g., which and how plasmids/ISs are involved in the evolution of HCH-degraders) remains unclear.

In this study, to gain further insight into the functional evolution of bacterial genomes, the complete genome sequences of three other \textsuperscript{c}-HCH-degraders, \textit{Sphingomonas} sp. MM-1,\textsuperscript{22,23} \textit{Sphingobium} sp. MI1205,\textsuperscript{24,25} and \textit{Sphingobium} sp. TKS,\textsuperscript{26} were determined. On the basis of our comparison of the complete genome sequences of these three strains and UT26, in association with several supporting experimental data, the evolution of \textsuperscript{c}-HCH-degrading bacterial strains is discussed.

2. Materials and methods

2.1. Bacterial strains, plasmids, and culture conditions

Strains and plasmids used in this study are listed in Supplementary Table S1. All sphingomonad (the collective name of \textit{Sphingomonas}, \textit{Sphingobium}, \textit{Novosphingobium}, \textit{Sphingopyxis}, and their related genera)\textsuperscript{27} strains were cultured at 30°C in 1/3LB medium or 1/10W minimal medium.\textsuperscript{24} If needed, antibiotics were added at the following concentrations: tetracycline (Tc) 20 mg/ml, nalidixic acid (Nal) 100 mg/ml, and kanamycin (Km) 50 mg/ml for UT26; Tc 5 mg/ml, Nal 100 mg/ml, gentamycin (Gm) 2 mg/ml, and Km 50 mg/ml for MM-1; Tc 2 mg/ml, Nal 100 mg/ml, and Km 50 mg/ml for TKS; and Tc 2 mg/ml, Nal 100 mg/ml, and Km 50 mg/ml for MI1205. \textit{Escherichia coli} DH5\textsuperscript{a} for genetic manipulation was grown at 37°C in LB.\textsuperscript{28} If needed, antibiotics were added at the following concentrations: Tc 20 mg/ml, Gm 20 mg/ml, and Km 50 mg/ml. The solid media were prepared by the addition of 1.5% agar.

2.2. Construction of plasmids and strains

The 3.6-kb region containing the MM-1 linKLMNb genes was amplified by PCR using the primer set of MM_linKLMN_F and MM_linKLMN_R (Supplementary Table S2), and the amplified fragment was cloned into a broad-host-range vector, pKS13P,\textsuperscript{29} under the \textit{linA} constitutive promoter\textsuperscript{30} to generate pKSR1020. The whole region of the linFb gene of strain TKS was amplified by PCR using the primer set of TKS_MAR_F_Hind and TKS_MAR_R_Bam (Fig. 1). The amplified fragment was cloned into the plasmid pKSR1020, and the recombinant plasmid was introduced into \textit{Sphingobium} sp. TKS by transformation. The resulting strain was designated as TKS_HAL.

Figure 1. Degradation pathway of \textsuperscript{c}-HCH in UT26. Compounds: 1, \textsuperscript{c}-HCH; 2, \textsuperscript{c}-pentachlorocyclohexene; 3, 1,3,4,6-tetrachloro-1,4-cyclohexadiene; 4, 1,2,4-trichlorobenzene; 5, 2,4,5-trichloro-2,5-cyclohexadiene-1-ol; 6, 2,5-dichlorophenol; 7, 2,5-dichloro-1,2,5-cyclohexadiene-1,4-diol; 8, 2,5-dichlorohydroquinone; 9, chlorohydroquinone; 10, acylchloride; 11, hydroquinone; 12, \textit{γ}-hydroxymuconic semialdehyde; 13, maleylacetate; 14, \textit{β}-ketoadipate; 15, 3-oxoadipyl-CoA; 16, succinyl-CoA; 17, acetyl-CoA. TCA, citrate/tricarboxylic acid cycle. Note that the compounds 4 and 6 are dead-end products. Spontaneous and non-enzymatic reaction is indicated by grey arrows. The reaction marked with ‘?’ has not been identified. Compounds in parentheses have not been directly detected.
(Supplementary Table S2), digested with BamHI and HindIII, and cloned into the corresponding sites of pBBR1-MCS2 to generate pBLFb. To disrupt the genomic linFb gene in TKS, an internal part of linFb was amplified by PCR using the primer set of TKS_MAR, single_F_Eco and TKS_MAR_single_R_Hind (Supplementary Table S2), digested with EcoRI and HindIII, and cloned into the corresponding sites of pEX18Gm. The resultant plasmid pEDLFb was introduced into TKS by electroporation to select the Gm-resistant transformants. Subsequent PCR analysis of one such transformant (TKSdLFb) confirmed that its genomic linFb gene was disrupted by reciprocal homologous recombination through the single crossover-mediated integration of pEDLFb (Supplementary Fig. S1).

2.3. DNA manipulations and Sanger sequencing

Established methods were employed for the preparation of plasmids and genomic DNAs, their digestion with restriction endonucleases, ligation, and agarose gel electrophoresis, and the transformation of E. coli cells. Electroporation of sphingomonad strains was performed as described previously. PCR for cloning was performed with KOD-Plus DNA polymerase (TOYOBO, Osaka, Japan). The primers used are listed in Supplementary Table S2. The Sanger sequencing was performed using an ABI PRISM 3130xl sequencer and ABI Prism Big Dye Terminator v3.1 Kit (Applied Biosystems).

2.4. Genome sequencing and annotation analyses

Fragment reads of genomic DNA of MM-1, MI1205, and TKS were obtained by the Roche 454 and Illumina HiSeq 2000 sequencing systems, and were assembled using the Newbler programme (Roche). The GenoFinisher and AceFileViewer programmes were used to finish the sequencing completion. More detailed information for these procedures have been published elsewhere. Sequencing gap regions were amplified by PCR using KOD FX (TOYOBO) or Ex Taq (TaKaRa) by using the total DNA of the respective strains as templates, and the resultant DNA fragments were sequenced using primers for PCR amplification. Pulsed-field-gel electrophoresis was also performed as described previously to support the whole genome sequencing data of MM1205 and TKS (data not shown). The annotation data of the complete genome sequences were obtained by PGAAP (Prokaryotic Genome Annotation Pipeline, http://www.ncbi.nlm.nih.gov/genome/annotation_prok/; 22 August 2016, date last accessed), and was curated with the dedicated software bundled in the GenomeMatcher programme (http://www.inei.tohoku.ac.jp/joho/gmProject/gmhome.html; 22 August 2016, date last accessed) as well as by consulting the MiGAP auto-annotation system (Microbial Genome Annotation Pipeline, http://www.migap.org/; 22 August 2016, date last accessed).

2.5. Computational analyses of sequence data

The nucleotide and protein sequences were analysed using the Genetyx programme version 13–18 (Genetyx Corp., Tokyo, Japan). Homology searches were performed using the BLAST programmes available at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/BLAST/; 22 August 2016, date last accessed) with the default parameters. Venn diagram was depicted using the result obtained by BLASTClust analysis (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/; 22 August 2016, date last accessed) for all predicted open reading frames (ORFs) of the four γ-HCH degraders with a parameter set ‘-p T -L .6 -b T -S 60 -a 24’. Comparative analyses of DNA sequences were performed by GenomeMatcher using BLASTN with a parameter set ‘-F F -W 21 -e 0.01’. Conserved motifs and repeat sequences were searched by GenomeMatcher. All neighbour-joining phylogenetic trees shown in this study were constructed using MAFFT programme (http://mafft.ebi.ac.uk/alignment/software/; 22 August 2016, date last accessed) and visualized by NJplot software (http://pbil.univ-lyon1.fr/software/njplot.html; 22 August 2016, date last accessed). Transposable elements were identified by analyses of regions containing putative transposase genes, i.e. mutual BLASTN analysis with a parameter set ‘-F F -W 21 -e 0.01’ and searches for inverted and direct repeats using Dot Match mode of GenomeMatcher. For detection of candidate ORFs relevant to metabolisms of aromatic compounds, all ORFs obtained from each genome were BLASTP-searched with parameters ‘-e 1e-5 -b 5 -F F’ and threshold identity ≥50%, query and reference sequence coverage ≥50% against an in-house database for the enzymes for degradation of aromatic compounds.

2.6. Entrapment of transposable elements

pGEN500, an entrapment plasmid vector of transposable elements, was introduced into UT26, MM-1, TKS, and MI1205 by electroporation or mating using E. coli S17-1, and the cells carrying pGEN500-containing plasmids were selected on the 1/3LB agar plate containing Tc. The pGEN500-containing strains were thereafter plated on 1/3LB agar containing Tc and 10% sucrose (w/v) at 30°C. The colonies formed on the plates were analysed for their resident pGEN500 derivatives. Such derivatives with enlarged sizes were postulated to be formed by the insertion of endogenous transposable elements in the sacB gene on pGEN500. The insertion event in each derivative was investigated by PCR using the multiple pairs of primers listed in Supplementary Table S2, and the insert was sequenced by the Sanger method.

3. Results and discussion

3.1. Complete genome sequences of three γ-HCH-degrading sphingomonad strains

Most of the aerobic γ-HCH-degrading bacterial strains that have been critically analysed at the genetic level are sphingomonads that belong to Alphaproteobacteria. We have previously published an article describing our detailed analysis of the UT26 genome. In this study, the complete genome sequences of three other γ-HCH-degrading strains, Sphingomonas sp. MM-1, Sphingobium sp. MI1205, and Sphingobium sp. TKS, were determined.

The basic genome organizations of the four γ-HCH-degrading strains isolated from different geographical areas are summarized in Table 1. Comparative analysis of the 16S rRNA genes indicated that these four γ-HCH-degrading strains are phylogenetically diverse among related sphingomonad strains (Fig. 2). All predicted ORFs of these four strains (19,312) were clustered into 10,325 ORF clusters. Figure 3 summarizes the result as a Venn diagram that shows the numbers of shared and unique ORF clusters among the four strains; these four strains each have 1,190–2,346 unique ORF clusters, but share only 1,288 ones. This observation supports the phylogenetic divergence of the four γ-HCH degraders. These results also strongly suggest that the four degraders independently acquired γ-HCH-degradation ability, and thus it is unlikely that the four strains have been derived from one ancestral γ-HCH-degrader.

In addition to the genes for γ-HCH degradation (see below), putative genes for the degradation of various aromatic compounds, toluenophenol, chlorophenol, anthranilate, and homogentisate, reside in the UT26 genome. These genes constitute four clusters for the
| Strain name | Isolated site | Source type | Replicon | Length (bp) | Number of | Acc no. | Replicon type | lin genes<sup>b</sup> | Reference for strain |
|------------|--------------|-------------|----------|-------------|-----------|---------|---------------|-------------------|---------------------|
| S. japonicum UT26 | Tokyo, Japan | Soil artificially polluted with γ-HCH | Chromosome 1 | 3,514,822 | 1 | 3,529 | AP010803 | Chr | linA, linB, linC, linKLNM | 18 |
|              |              |             | pCHQ1    | 190,974  | 0 | 224   | AP010805 | pCHQ1 | linRED |              |
|              |              |             | pUT1     | 31,776   | 0 | 44    | AP010806 | pUT1 | — |              |
|              |              |             | pUT2     | 5,398    | 0 | 8     | AP010807 | pUT2 | — |              |
|              |              |             | total    | 4,424,862 | 3 | 4,394 | 13 |
| Sphingomonas sp. MM-1 | Lucknow, India | Soil polluted with HCH isomers | Chromosome | 4,054,833 | 2 | 3,801 | CP004036 | Chr | linKbLbMbNb | 22 |
|              |              |             | pISP0    | 275,840  | 0 | 251   | AP010437 | pCHQ1 | linF, linGHII |              |
|              |              |             | pISP1<sup>c</sup> | 172,140 | 0 | 174   | AP010438 | UT26_Chrom 2/pISP4 | linA, linC, linF' |              |
|              |              |             | pISP2    | 53,841   | 0 | 52    | AP010439 | pUT1 | — |              |
|              |              |             | pISP3    | 43,776   | 0 | 44    | AP010440 | pISP3 | linRED |              |
|              |              |             | pISP4    | 33,183   | 0 | 39    | AP010441 | pISP4 | linB, linC, linF' |              |
|              |              |             | total    | 4,633,613 | 2 | 4,361 | 15 |
| Sphingobium sp. TKS | Kyushu, Japan | Sediment polluted with HCH isomers | Chromosome 1<sup>c</sup> | 4,249,857 | 1 | 4,127 | CP005083 | Chr/UT26_Chrom 2 | linB, linC, linF', linFB, linKLNM | This study |
|              |              |             | pTK1<sup>c</sup> | 989,120 | 2 | 843   | AP010584 | UT26_Chrom 2 | linGHII_homologue |              |
|              |              |             | pTK2      | 520,614  | 0 | 470   | AP010585 | UT26_Chrom 2/ pCHQ1 | — |              |
|              |              |             | pTK3<sup>c</sup> | 195,308 | 0 | 182   | CP005086 | pTK2 | — |              |
|              |              |             | pTK4<sup>c</sup> | 87,635  | 0 | 92    | CP005087 | pISP4/pTK3_type 1/pTK3_type 2 | linB, linC, linF' |              |
|              |              |             | pTK5      | 75,938   | 0 | 86    | CP005088 | pUT1/pISP4 | linA, linC |              |
|              |              |             | pTK6      | 53,908   | 0 | 76    | CP005089 | pLB1 | — |              |
|              |              |             | pTK7      | 34,300   | 0 | 35    | CP005090 | pISP3 | linRED |              |
|              |              |             | pTK8      | 9,585    | 0 | 12    | CP005091 | pTK7 | — |              |
|              |              |             | pTK9      | 7,223    | 0 | 11    | CP005092 | pTK8 | — |              |
|              |              |             | total    | 6,228,879 | 3 | 5,987 | 29 |
| Sphingobium sp. MI1205 | Miyagi, Japan | Soil polluted with HCH isomers | Chromosome 1 | 3,351,250 | 1 | 3,285 | AP005188 | Chr | linKLNM | 24 |
|              |              |             | pMI1      | 3,351,250 | 1 | 3,285 | AP005188 | Chr | linKLNM |              |
|              |              |             | pMI2<sup>c</sup> | 567,154 | 1 | 516   | AP005189 | UT26_Chrom 2 | — |              |
|              |              |             | pMI3<sup>c</sup> | 292,135 | 0 | 299   | AP005190 | UT26_Chrom 2 | — |              |
|              |              |             | pMI4<sup>c</sup> | 287,488 | 0 | 323   | AP005191 | pUT1/UT26_Chrom 2 | linA, linRD |              |
|              |              |             | total    | 4,619,375 | 2 | 4,570 | 24 |

<sup>a</sup>See Table 3

<sup>b</sup>linF' and linF'' are probably pseudogenes (see Supplementary Fig. S2).

<sup>c</sup>Replicons having more than one rep genes.
degradation of the respective compounds, and each cluster contains all the genes necessary for the conversion of each compound to the metabolites in the central metabolic pathway, strongly suggesting that UT26 is able to utilize these compounds. Similarly, several putative genes for the degradation of aromatic compounds were found in the MM-1, MI1205, and TKS genomes (Supplementary Table S3). The potential of the four γ-HCH degraders for the degradation of aromatic compounds was estimated more comprehensively by BLASTP search of all their ORFs against our previously constructed in-house database which consists of enzymes for the degradation of aromatic compounds,41 and the result was summarized in Supplementary Table S4. The numbers of ORFs potentially involved in the degradation of aromatic compounds in the four strains (62, 46, 27, and 25 for TKS, UT26, MI1205, and MM-1, respectively) are much smaller than those in versatile recalcitrant pollutant degraders, Cupriavidus necator JMP134,43,44 and Burkholderia xenovorans LB40045,46 (149 and 135 for JMP134 and LB400, respectively). Especially, those in UT26, MI1205, and MM-1 are even smaller than those in typical metabolically versatile soil bacterial strains Burkholderia multivorans ATCC 1761647–50 and Pseudomonas putida KT244051 (73 and 62 for KT2440 and ATCC 17616, respectively). These results indicate that our sphingomonad strains are ‘specialists’ for γ-HCH degradation, but not ‘generalists’ for the degradation of many recalcitrant compounds.

3.2. The lin genes for γ-HCH utilization

UT26 converts γ-HCH to β-ketoadipate via reactions catalysed by dehydrochlorinase (LinA), haloalkane dehalogenase (LinB), dehydrogenase (LinC), reductive dechlorinase (LinD), ring-cleavage dioxygenase (LinE), and maleylacetate reductase (MAR) (LinF); β-ketoadipate is thereafter converted to succinyl-CoA and acetyl-CoA.
by succinyl-CoA:3-oxoacid CoA transferase (LinGH) and β-ketoacyl CoA thiolase (LinJ), respectively (Fig. 1).16,18 In addition to genes for these catabolic enzymes and their regulatory genes (linK for linDE and linC for linGHJ),18,52 the linKLNM genes encoding a putative ABC-transporter system are necessary for the γ-HCH utilization in UT26.29 The linA, linB, linC, and linF genes, and the linRED, linGHIJ, and linKLNM clusters are dispersed on the UT26 genome.18 Since the β-ketoacid pathway is often used by environmental bacterial strains,53 the lin genes for the conversion of γ-HCH to β-ketoacid (linA to linF) are peculiar to the γ-HCH-degrading pathway. In particular, the linA gene is unique because it does not show significant similarity to any sequences in the databases except for the almost identical (>90% identity) linA genes from other bacterial strains and metagenomes.7,16

The MM-1, MI1205, and TKS genomes carry linA, linB, and linC genes and a linRED cluster that are almost identical (>98% identity at the DNA level) to those of UT26 (Table 2). The former two strains additionally carry a linF gene and linGHJ cluster that are almost identical (>98% identity at the DNA level) to those in UT26, strongly suggesting that γ-HCH is degraded in these two strains by the same pathway as in UT26 (Fig. 1). The last strain lacks the linFUT26 and linGHIJUT26 cluster for maleylacetate metabolism (Fig. 1). Although two copies of the truncated version of linF (named linF′) (Supplementary Fig. S2) are present in the TKS genome, linF′ was assumed not to encode functional MAR, since linF′ misses more than one half of the intact linF gene (Supplementary Fig. S2). TKS instead carries another putative gene, designated linFb, on Chr1. Although LinFb showed only 49% identity to LinFUT26 (Table 2), it showed much higher similarity with other known MAR proteins, such as TidF from Bordetella pertussis,54 Achromobacter denitrificans,55 Burkholderia sp. M701 (Acc. no. YP_008864525.), and Comamonas testosteroni66 (83.78%, 78%, and 71%, respectively). Interestingly, Sphingobium sp. HDIPO4, a recently isolated HCH degrader, has the identical linFb gene, although its start codon is annotated at a position different from that in TKS.37 To clarify the LinFb function, its gene in TKS was disrupted (Supplementary Fig. S1). The resultant strain did not grow on a minimal agar plate supplemented with γ-HCH as a sole source of carbon and energy, and this growth defect was reversed by the supply of the intact linFb gene (Supplementary Fig. S3). In addition, the γ-HCH utilization defect of UT1023d, a linF′ mutant of UT26,34 was reversed by the supply of the linFb gene (data not shown). These results clearly demonstrated that the linFb encodes a MAR that is functional for the γ-HCH utilization. Although the linGHIJUT26 cluster was not found in the TKS genome, many homologues of linG, linH, and linI were found (Supplementary Table S3). Among these homologues, one set of linGH homologues is located just downstream of linFb on Chr1TKS (Supplementary Fig. S4A) and only one set of linGHIJ homologues exists as a cluster on Chr2TKS (Supplementary Fig. S4B). Some of these homologues may be functional for the β-ketoacid metabolism, and this experimental confirmation is necessary. These results strongly suggested that γ-HCH is also degraded in TKS by the same pathway as in UT26 (Fig. 1).

The linKLMNUT26 homologues have been found in various bacterial strains16,22 and were also found in the MM-1, MI1205, and TKS genomes (Table 2). However, their similarities to linKLMNUT26 are lower (<93% identity at DNA level; Table 2) than in the case of the other lin genes mentioned earlier, and this divergence reflects the phylogenetic relationship of their hosts (Fig. 2 and Supplementary Fig. S5), suggesting that the linKLNM system is one of the inherent functions necessary for γ-HCH utilization in sphingomonads. In particular, the linKLNM homologues of MM-1, which is phylogenetically the most distant strain from UT26 (Fig. 2), show a relatively low similarity with linKLMNUT26 (42–75% identities at the amino acid level; Table 2 and Supplementary Fig. S5), and they were designated linKbLbMbNb. To confirm their function for γ-HCH utilization, we attempted to disrupt the linKbLbMbNb gene cluster in MM-1. However, such an expected disruptant could not be constructed because unknown DNA rearrangements often occurred in MM-1. As an alternative confirmation, a plasmid containing the linKbLbMbNbM1 genome cluster was introduced into RE1, a linKLNMUT26 disruptant of UT26.29 The growth of RE1 in LB medium was inhibited by the addition of γ-HCH (Supplementary Fig. S6A),29 and this inhibition was suppressed by the supply of the linKbLbMbNb cluster (Supplementary Fig. S6B) as well as the linKLNMUT26 one (Supplementary Fig. S6C),29 strongly suggesting that the both clusters function in the same way for the γ-HCH utilization. These results supported our hypothesis that the linKLNM system is one of the inherent functions necessary for γ-HCH utilization in sphingomonads. However, the possibility cannot be excluded that other functional homologue(s) of linKLNM system exist in MM-1.

All of the four strains carry almost identical linA to linF genes (designated ‘specific’ lin genes) (Table 2), suggesting they acquired such genes by lateral gene transfer. However, the specific lin genes are dispersed on multiple replicons in the four strains (Table 2). In UT26, linA to linC are located on Chr1, and only the linRED cluster is located on a plasmid. On the other hands, all the specific lin genes are dispersed on multiple plasmids with various combinations in other three strains, although additional copies of linB and linC are also located on Chr1 in TKS (Table 1). Furthermore, replicon types of such plasmids carrying the specific lin genes are various (Table 3). These observations indicate that these four strains did not simply acquire all the specific lin genes at once as a cluster. This contrasts with other aromatic compound-degrading strains, which can acquire a whole set of responsible genes by the conjugative transfer of plasmids and/or integrative and conjugative elements.58–60

Although, in the present study, we only described the overall genetic repertoire of the lin genes for γ-HCH utilization, the composition of genes for LinA and LinB variants and their copy numbers and expression levels are important for the degradation performance of host strains toward HCH isomers, since the LinA and LinB variants show different levels of enzymatic activity toward different HCH isomers and their metabolites.72,16,61–64 However, in order to properly discuss this point from a genomic viewpoint, additional fundamental biochemical and experimental data will be needed.

3.3. Replication/partition-encoding regions of plasmids and plasmid-type replicons in sphingomonad strains

The putative replication origins (oriCs) of the main chromosomes (Chr1s) of TKS, MI1205, and MM-1 were, as in the case with Chr1UT26,18 found to be of alphaproteobacterial-chromosome-type63,66 these oriCs were located upstream of the uroporphyrinogen decarboxylase gene (hemeE) with multiple DNA boxes (TT[AT/TA]NCACA) (Supplementary Fig. S7).67 On the other hand, as in the case of Chr2UT26, both Chr2TKS and Chr2MI1205 have the plasmid-type replication and active partition systems.28 These three plasmid-type chromosomes and the 21 plasmids in our four strains and the plasmids from other sphingomonads were, on the basis of the similarities of their RepA (DNA replication initiator) proteins, classified into 13 types (Table 3). Although the importance of
| Gene  | Function                  | UT26        | MM-1        | TKS         | MI1205      |
|-------|---------------------------|-------------|-------------|-------------|-------------|
| linA  | Dehydrochlorinase         | 156 Chr1_1860686-1861156_c | 156 pBS1_135477_14017    | 98 (153/156) | 156 pTK4_18583_19053    | 97 (290/296) |
| linB  | Halodioxynuclease         | 296 Chr1_1966541-1967631 | 296 pBS1_18987_19484_c   | 98 (283/286) | 296 pTK3_34952_35842    | 97 (290/296) |
| linC  | Dehydrogenase             | 250 Chr1_566609-567361_c | 250 pISP1_147367_148119  | 99 (249/250) | 250 pMI2_260065_260535_c | identical |
| linD  | Reductive dechlorinase    | 346 pCHR1_110947-111926_c | 346 pISP3_15908_16948    | identical   | 346 pMI1_148490_149500_c | identical |
| linE  | Ring-cleavage dioxygenase | 321 pCHR1_114235-115200_c | 321 pISP3_12695_13660    | similar     | 321 pMI1_151777_152742_c | identical |
| linF  | Maleylactate reductase    | 352 Chr2_562332-563390_c | 352 pISP0_110260_111318_c | identical   | 352 pMI1_152874_153785_c | identical |
| linG  | Acyl-CoA transferase, alpha subunit | 215 Chr2_603108-603755 | 215 pISP0_152221_152868 | identical   | 215 pMI1_152874_153785_c | identical |
| linH  | Acyl-CoA transferase, beta subunit | 212 Chr2_603755-604393 | 212 pISP0_152868_153506 | identical   | 212 pMI1_152874_153785_c | identical |
| linI  | IclR-family transcriptional regulator | 267 Chr2_601921-602121_c | 267 pISP0_151281_152084_c | 99 (913/916) | 267 pMI1_152874_153785_c | identical |
| linJ  | Thiolase                  | 403 Chr2_600921-601321_c | 403 pISP0_150034_151245_c | 99 (122/122) | 403 pMI1_152874_153785_c | identical |
| linK  | Putative ABC transporter system, inner membrane protein | 376 Chr1_193477-20477 | 376 pCHR1_235484_235952_c | 99 (122/122) | 376 pCHR1_235484_235952_c | 99 (122/122) |
| linL  | Putative ABC transporter system, ATPase | 282 Chr1_210477-211235 | 282 pCHR1_235484_235952_c | 99 (122/122) | 282 pCHR1_235484_235952_c | 99 (122/122) |
| linM  | Putative ABC transporter system, periplasmic protein | 320 Chr1_21329-22291 | 320 pCHR1_235484_235952_c | 99 (122/122) | 320 pCHR1_235484_235952_c | 99 (122/122) |
| linN  | Putative ABC transporter system, lipoprotein | 202 Chr1_22299-22397 | 202 pCHR1_235484_235952_c | 99 (122/122) | 202 pCHR1_235484_235952_c | 99 (122/122) |

| Gene  | Function                  | UT26        | MM-1        | TKS         | MI1205      |
|-------|---------------------------|-------------|-------------|-------------|-------------|
| linA  | Dehydrochlorinase         | 156 Chr1_1860686-1861156_c | 156 pBS1_135477_14017    | 98 (153/156) | 156 pTK4_18583_19053    | 97 (290/296) |
| linB  | Halodioxynuclease         | 296 Chr1_1966541-1967631 | 296 pBS1_18987_19484_c   | 98 (283/286) | 296 pTK3_34952_35842    | 97 (290/296) |
| linC  | Dehydrogenase             | 250 Chr1_566609-567361_c | 250 pISP1_147367_148119  | 99 (249/250) | 250 pMI2_260065_260535_c | identical |
| linD  | Reductive dechlorinase    | 346 pCHR1_110947-111926_c | 346 pISP3_15908_16948    | identical   | 346 pMI1_152874_153785_c | identical |
| linE  | Ring-cleavage dioxygenase | 321 pCHR1_114235-115200_c | 321 pISP3_12695_13660    | similar     | 321 pMI1_152874_153785_c | identical |
| linF  | Maleylactate reductase    | 352 Chr2_562332-563390_c | 352 pISP0_110260_111318_c | identical   | 352 pMI1_152874_153785_c | identical |
| linF' | Putative ABC transporter system, periplasmic protein | 316 Chr1_193477-20477 | 316 pCHR1_235484_235952_c | 99 (122/122) | 316 pCHR1_235484_235952_c | 99 (122/122) |
| linF0 | Putative ABC transporter system, inner membrane protein | 376 Chr1_193477-20477 | 376 pCHR1_235484_235952_c | 99 (122/122) | 376 pCHR1_235484_235952_c | 99 (122/122) |

*a, encoding on complementary strand.
*b, linF and linF' are probably pseudogenes.
'c, linEb is homologue of PcpA (96% identity) and partially involved in γ-HCH degradation in UT26.34
| Replicon | Site (bp) | Host | Host feature | RepA protein | Type | Gene cluster for degradation | Genes for degradation | Accession number | Reference |
|----------|----------|------|--------------|--------------|------|------------------------------|----------------------|------------------|-----------|
| UT26_Chr2 | 681,892 | S. japonicum UT26 | γ-HCH degradation | 1_1230 | 399 | repABC | linF, linGHIJ | CP003588 | 18 |
| L-1_Chr2 | 1,368,670 | Sphingobium chlorophosphocolum L-1 | PCP degradation | 321340_322506 | 396 | repABC | linF, linEMAC | CP002799 | 11 |
| TKS_Chr2 | 989,120 | Sphingobium sp. TKS | γ-HCH degradation | 1_1206 | 401 | repABC | linF, linGHIJ | CP005084 | 26 |
| Mi1_Chr2 | 567,154 | Sphingobium sp. Mi1205 | γ-HCH degradation | 1_1209 | 403 | repABC | linF, linGHIJ | CP005189 | 15 |
| pNL1 | 487,268 | Novosphingobium aromaticivorans DSM 12444 | aromatic compounds degradation | 209493_210644 | 401 | repABC | linF, linGHIJ | CP006677 | unpublished |
| lpl | 192,103 | Novosphingobium sp. PP1Y | aromatic compounds degradation | 88922_90199 | 425 | repABC | linF, linGHIJ | FR568640 | 80 |
| plsp1b | 172,140 | Sphingomonas sp. MM-1 | γ-HCH degradation | 114750_115880 | 376 | repABC | F, linA, linC, linF | CP004018 | 23 |
| TKS_Chrb | 4,249,837 | Sphingobium sp. TKS | γ-HCH degradation | 252492_253622 | 376 | repABC | linA, linC, linF | CP005083 | 26 |
| pMI2 | 287,488 | Sphingobium sp. Mi1205 | γ-HCH degradation | 183535_184485 | 376 | repABC | F, linA, linC | CP005191 | 25 |
| pMI1 | 292,135 | Sphingobium sp. Mi1205 | γ-HCH degradation | 1_1224 | 401 | repABC | F, linA, linC, linRED, linE, linF, linGHIJ | CP005190 | 25 |
| pTK1b | 520,614 | Sphingobium sp. TKS | γ-HCH degradation | 312340_313539 | 399 | repABC | F, Ti | CP005085 | 26 |
| pCHQ1 | 190,794 | S. japonicum UT26 | γ-HCH degradation | 1_1164 | 387 | — | — | CP003585 | 26 |
| pTK1b | 520,614 | Sphingobium sp. TKS | γ-HCH degradation | 1_1164 | 387 | — | — | CP003585 | 26 |
| pS1GP | 148,001 | Sphingobium sp. SYK-4 | linF degradation | 1_1316 | 387 | — | — | AP012223 | 81 |
| pSPhIC101 | 123,733 | Sphingobium chlorophosphocolum L-1 | PCP degradation | 4083_48171 | 362 | — | — | CP002800 | 11 |
| pSP0 | 275,840 | Sphingobium sp. MM-1 | γ-HCH degradation | 1_1104 | 367 | — | — | CP004017 | 23 |
| plUT1 | 317,766 | S. japonicum UT26 | γ-HCH degradation | 1_1104 | 367 | — | — | CP003585 | 26 |
| pMI2b | 287,488 | Sphingobium sp. Mi1205 | γ-HCH degradation | 1_1104 | 367 | — | — | CP003585 | 26 |
| pSP2 | 53,841 | Sphingobium sp. MM-1 | γ-HCH degradation | 1_1104 | 367 | — | — | CP003585 | 26 |
| pTK4b | 75,938 | Sphingobium sp. TKS | γ-HCH degradation | 1_1104 | 367 | — | — | CP003585 | 26 |
| pSP3 | 43,776 | Sphingobium sp. MM-1 | γ-HCH degradation | 37217_38329 | 370 | — | — | CP004040 | 23 |
| pMHo | 32,974 | Sphingobium sp. Mi1205 | γ-HCH degradation | 1_1113 | 370 | — | — | CP005193 | 25 |
| pTK6 | 34,300 | Sphingobium sp. TKS | γ-HCH degradation | 1_1113 | 370 | — | — | CP005090 | 26 |
| pTK1b | 87,635 | Sphingobium sp. TKS | γ-HCH degradation | 1_960 | 319 | — | — | CP005087 | 26 |
| pTK2b | 87,635 | Sphingobium sp. TKS | γ-HCH degradation | 52874_53823 | 315 | — | — | CP005087 | 26 |
| pSP4 | 33,183 | Sphingobium sp. MM-1 | γ-HCH degradation | 101_1012 | 303 | — | — | CP004041 | 23 |
| pSP1b | 172,140 | Sphingobium sp. MM-1 | γ-HCH degradation | 158477_159388 | 303 | — | — | CP004038 | 23 |
| pTK3b | 87,635 | Sphingobium sp. MM-1 | γ-HCH degradation | 192631_201754 | 303 | — | — | CP005193 | 25 |
| pTK4b | 75,938 | Sphingobium sp. TKS | γ-HCH degradation | 56463_57374 | 303 | — | — | CP005088 | 26 |
| pMIb | 88,374 | Sphingobium sp. Mi1205 | γ-HCH degradation | 30095_31006 | 303 | — | — | CP005192 | 25 |
| pMHo | 32,974 | Sphingobium sp. Mi1205 | γ-HCH degradation | 18420_19331 | 303 | — | — | CP005193 | 25 |
| pTK2 | 193,308 | Sphingobium sp. TKS | γ-HCH degradation | 1_1212 | 373 | — | — | CP005086 | 26 |
| pTK7 | 9,585 | Sphingobium sp. TKS | γ-HCH degradation | 1_1107 | 368 | — | — | CP005091 | 26 |
| pTK8 | 7,223 | Sphingobium sp. TKS | γ-HCH degradation | 1_1032 | 343 | — | — | CP005092 | 26 |
| pNL1 | 184,462 | Novosphingobium aromaticivorans DSM 12444 | aromatic compounds degradation | 86362_87666 | 434 | — | — | CP006676 | unpublished |
| pCAR3 | 234,797 | Novosphingobium sp. KA1 | carboxyclic degradation | 200394_201898 | 434 | 91 | — | AB270330 | 82 |
| pS1WIT02 | 222,757 | Sphingobium wittichii RW1 | aromatic compounds degradation | 63589_64893 | 434 | 82 | — | FR456641 | 80 |
| plB1 | 65,998 | unidentified soil bacterium (S. japonicum UT26) | γ-HCH degradation | 1_783 | 260 | — | Ti | AB44976 | 74 |
| pMIb | 88,374 | Sphingobium sp. Mi1205 | γ-HCH degradation | 1_783 | 260 | — | — | CP005192 | 25 |
| plA2 | 62,341 | Novosphingobium pontiaceticum U6E-1 | γ-HCH degradation | 40543_41325 | 260 | 98 | — | CP005089 | 26 |
| pTK5 | 53,908 | Sphingobium sp. TKS | γ-HCH degradation | 3_783 | 260 | 97 | — | CP005089 | 26 |
| pTK2 | 5,398 | S. japonicum UT26 | γ-HCH degradation | 1_1203 | 400 | — | — | AP010807 | 18 |
| pTK9 | 5,391 | Sphingobium sp. TKS | γ-HCH degradation | 1_654 | 217 | — | — | CP005093 | 26 |

*The representative replicons of each type ones are indicated in bold in the first column.

Replions having more than one rep genes.

Start codon is differently annotated for the same DNA region.

Identical with each other.
plasmids in sphingomonads has been recognized, no detailed analysis of their fundamental machineries was reported. In addition, RepA proteins of plasmids in sphingomonads show a very low level of similarity to those of well-studied plasmids (e.g. IncP-1, F, IncP-7, and IncP-9 plasmids), and thus we compared only plasmids in sphingomonads in this study. Since the RepA proteins of the 13 types are very divergent, they were further categorized into three major groups, in each of which the RepA proteins exhibit 22–60% identity: (i) the Chr2UT26- and pCHQ1-types (Fig. 4A), (ii) the pUT1-, pISP3-, pTK3_1-, and pTK3_2-types (Fig. 4B), and (iii) the pISP4-, pTK2-, pTK7-, and pTK8-types (Fig. 4C). Based on our BLASTP analysis, the RepA proteins of pNL1, pLB1, and pUT2 did not show similarity to those of any of the other types of plasmids listed in Table 3, although the RepA of pUT2 was similar to those of the IncP-9 family of plasmids.

Figure 4D schematically shows the organizations of the repA-flanking regions in the 12 representative plasmids (note that pTK3 has three types of repA genes), and many of these regions also carry the putative replication origin (oriV) sequences as well as the putative genes for the active partition systems, each with the putative parS (cis-acting centromeric) sequences, the parB gene encoding the parS-binding protein, and the parA gene encoding the NTPase that is capable of binding the parS-ParB complex. The Chr2UT26- and pCHQ1-type plasmids belong to the repABC-type plasmids, and putative palindromic parS sequences were found (Fig. 4D and Supplementary Table S6). The Chr2UT26-type plasmids have the parA-parB-par cluster, and the order of these three genes is conserved in other typical repABC-type plasmids, although the RepA proteins from the Chr2UT26-type plasmids are divergent in their sizes and similarities (Table 3 and Fig. 4A). In contrast, the pCHQ1-type plasmids have a repA-parA-par cluster (Fig. 4D) with nearly identical RepA proteins (Table 3). Other types of plasmids except the pLB1-type were categorized as iteron-type plasmids (Table 3), and direct repeats (iteron), DnaA box, and parS sequences were found in their repA-flanking regions (Fig. 4D and Supplementary Table S6). Each of the pISP1, Chr1TKS, pTK1, pTK3, pTK4, pML2, pML3, and pMI4 replicons appears to carry at least two repA genes, which are of different types (Tables 1 and 3). This observation suggested the frequent occurrence of fusions of ancestral plasmids (see below). Similar mosaic replicons carrying more than one repA gene have been reported in various bacterial strains.

Interestingly, six pISP4-type plasmids carry identical repA and parA genes, and five of them also have other types of repA genes (Table 3),

Figure 4. Phylogenetic trees of putative RepA proteins of Chr2UT26- and pCHQ1- (A), pUT1- (B), and pISP4- (C) types plasmids and organizations of repA-flanking regions of 12 representative sphingomonad plasmids (D). Classification of 13 types of plasmids and information on RepA protein sequences are summarized in Table 3. Neighbour-joining phylogenetic trees of the conserved sites, 260 aa (A), 262 aa (B), and 257 aa (C), respectively, were constructed. Bootstrap values calculated from 1,000 resampling using neighbour-joining are shown at the respective nodes. Length of lines reflects relative evolutionary distances among the sequences. RepA proteins of the representatives of the plasmid types (Table 3) are bolded. In panel D, the repA-flanking regions of plasmids whose putative RepA proteins show significant similarity are boxed. Pentagons indicate size and direction of ORF. Putative ORFs involved in replication and partition are filled with dark and light gray, respectively. Putative parS (palindromic TTNCACGTAA) and DnaA box (TTA or TTNACA) sequences are shown in red bars and red diamonds, respectively. Inverted repeats and repeat sequences are shown in blue and green bars, respectively. See Supplementary Table S6 for their sequences.
suggesting a prevalent fusion event of replicons in the pISP4-type plasmids (see below). It is noteworthy that all six pISP4-type plasmids contain the lin genes (Table 3), indicating that this type of plasmid plays an important role in dissemination of the lin genes.

3.4. Highly conserved regions of replicons in sphingomonad strains
Although multiple members in each of the pCHQ1-, pUT1-, pLB1-, pISP3-, pUT2-, and pISP4-type plasmids have almost identical repA-containing regions, the sizes and gene contents of the plasmid members in each type are diverse (Table 3). Therefore, we compared the overall structures in the six types of plasmids (Fig. 5). The 6.4-kb repA-parA-parB-containing region and the 10-kb repA-parA-containing region are conserved in all members of the pCHQ1- and pUT1-type plasmids, respectively (Fig. 5AB). The 25.5-kb region containing the pLB1-type repA and parB genes is conserved in pMI3 and pTK5 (Fig. 5C). Most of this region is also conserved in pLA2, although the region is divided into two parts and the parB gene is lacking (Fig. 5C). pMI3 is a fusion plasmid of the pLB1- and pISP4-type plasmids because the 9.1-kb repA-parA-containing region commonly conserved in the pISP4-type plasmids (Fig. 5F) is, together with the linC- and linF-containing region, inserted into the continuous region on pLB1 (Fig. 5C). A part of the 9.1-kb conserved region is also present in pTK4 (Fig. 5B). The 4.5-kb region containing the plSP3-type repA and parA genes is conserved in pTK6 and pMI4 (Fig. 5D). Two 1,645-bp plasmids, pUT2 and pTK9, differ by only 9 bp (Fig. 5E). As mentioned above, all the pISP4-type plasmids except for archetypal pISP4 have, in addition to the common repA gene, other distinct repA genes (Table 3), and probably are fusion plasmids. IS6100 or the Tn3-type transposon is located at the junctions of highly conserved regions of plSP4-type plasmids (Fig. 5F), suggesting that the plSP4-type plasmids are easily fused with other replicons via the transposition of IS6100 and/or the Tn3-type transposon. The repA, parA, and parB genes of the Chr2UT26-type replication machineries were also located on Chr1_TKS, pLSP1, pMI2, Chr2_TKS, pTK1, Chr2_MI1205, and pMI1 (Table 3), and the former
Figure 5. Structures of plasmids which have the highly conserved regions in the four γ-HCH-degrading sphingomonad strains. The highly conserved regions in pCHQ1-(A), pUT1-(B), pLB1-(C), pISP3-(D), pUT2-(E), and pISP4-(F) types plasmids are schematically shown. See Table 3 for the classification of plasmids. ORFs shown by pentagons are coloured as follows: red, rep and par genes; green, transposase gene of IS6100; cyan, putative genes for conjugal transfer; yellow, lin genes; and gray with gradient, transposition-related genes in other putative transposons. Almost identical regions are shown by the same background colours.
3.5. Genes for conjugal transfer of plasmids in HCH-degrading sphingomonads

The genes for conjugal transfer consist of those encoding proteins involved in mating pair formation (MpI) and DNA transfer and replication (Dtr). The mpI genes encode proteins that assemble in a large macromolecular structure called the Type IV secretion system (T4SS), whereas the dtr genes encode proteins that bind to the DNA at the origin of transfer region, oriT, forming a structure called a relaxosome. This modular gene organization is shared by most conjugative systems, showing a high degree of gene synteny conservation. Among the sphingomonad plasmids listed in Table 3, conjugal transferability of pCHQ1 and pLB1 has been experimentally confirmed, and these two plasmids have putative gene clusters for conjugal transfer similar to the vir gene cluster of Agrobacterium tumefaciens Ti plasmid, consisting of genes for MpI (VirB1 to VirB11) and Dtr (relaxase VirD2 and coupling protein VirD4) (Supplementary Fig. S9A). Putative vir gene clusters were also found on pLSp0, pTK1, pTK2, pTK5, and pMI3 (Supplementary Fig. S9A), indicating the potential self-transferability of these plasmids. However, the level of similarity of each component to the counterpart of Ti plasmid is relatively low, and only the phylogenetic relationship of the putative cytoplasmic ATPase component (VirB4), which is the essential and most conserved component of T4SS, encoded by these clusters is shown (Supplementary Fig. S9B).

pCHQ1 has another putative gene cluster for conjugal transfer similar to the tra gene cluster of F plasmid, and gene clusters similar to the tra gene cluster were also found on Chr1 TKS, pM11, pM12, pTK1, and pLSp1 (Supplementary Fig. S9C). As in the case with the gene clusters homologous to the vir gene cluster, the level of similarity of each component to its counterpart in F plasmid is relatively low, and only phylogenetic relationship of cytoplasmic ATPase component (TraC: VirB4 homologue in function) encoded by these clusters is shown (Supplementary Fig. S9D). However, traI and traD, which encode the relaxase and coupling protein, respectively, were not found on pCHQ1, and traD and traG, which encode the coupling protein and inner membrane platform component, respectively, are missing on pLSp1 (Supplementary Fig. S9C). Further experimental confirmation is necessary to demonstrate the self-transferability of these plasmids having putative gene clusters for conjugal transfer.

3.6. Transposable elements in four γ-HCH degraders

Many putative transposable elements including IS elements and Tn3-type transposons were found in the genomes of the four γ-HCH degraders (Supplementary Table S7). Although most of the IS elements are present as a single-copy form in the four strains (Supplementary Table S7), IS6100 is, as in the case of UT26, the most abundant element in the MM-1, TKS, and MI1205 genomes (15, 29, and 24 copies, respectively) (Table 1 and Supplementary Table S7). This suggests that IS6100 can transpose and increase its copy number in these γ-HCH degraders. To investigate the transposition activity of IS6100 and other transposable elements, the IS entrapment methodology using pGEN500 was applied for the four γ-HCH degraders. We conducted several independent analyses for each strain, and detected the successful transposition of IS6100, Issp1, ISSj02, ISSj12, and Tn6134 in UT26, IS6100, ISSj02, BSTik12, Tn6268, and Tn6269 in TKS, and IS6100, Issp1, ISMs02, ISMs08, Tn6137, and Tn6274 in MI1205 (Supplementary Table S7). On the other hand, this IS-entrapment system did not work well in MM-1 because of the high-frequency generation of the spontaneous sucrose-resistant mutants without the insertion of transposable elements into the sacB gene on pGEN500 (data not shown).

3.7. Inference of the past genome rearrangements via IS6100

IS6100 is often located in close proximity to the lin genes in the HCH-degrading strains and the metagenomic sequences from HCH-contaminated sites. IS6100 with a size of 880 bp is a ‘replicative’
IS element (Supplementary Fig. S10), and its transposition without apparent preference of target specificity causes the duplication of IS6100 with an 8-bp duplication of the target sequence. Therefore, the IS6100 transposition can generate three types of DNA rearrangements (Supplementary Fig. S11): intra-molecular transposition with a deletion/resolution (intra-replicon 1) or inversion (intra-replicon 2) event, and inter-molecular transposition with a fusion (inter-replicon) event. Our comparison of the regions just upstream and downstream of the 13 copies of IS6100 present on Chr1, Chr2, pCHQ1, and pUT1 of UT26 revealed five pairs of 8-bp sequences (Supplementary Table S8). On the basis of the IS6100 transposition mechanism (Supplementary Fig. S11), the most plausible past events caused by transposition of IS6100 can be inferred (Fig. 6A); it is indicated that not only simple transposition with inversion but also transposition accompanied with the fusion and resolution of replicons must have occurred. In a similar manner, seven, one, and four pairs of 8-bp sequences were found just upstream or downstream of IS6100 in TKS, MM-1, and MI1205, respectively (Supplementary Table S8), and the plausible past events mediated by transposition of IS6100 in these strains are depicted in Figure 6B–D.
The specific lin-flanking regions in the four strains were compared (Fig. 7). Not only the lin genes themselves (Table 2) but also their flanking regions are highly conserved (Fig. 7). Interestingly, such conserved regions are located very close to IS6100 and the distances between the IS6100 copies and the lin genes are varied (Fig. 7). This means that IS6100 is likely to play a crucial ‘editing’ role in the ‘trimming’ of ‘unnecessary regions’ for HCH utilization and the ‘gathering’ of the specific lin genes. At least, it is the most plausible that the transposition of IS6100 led to the diversification of the distribution and organization of the lin genes in the genomes. The distance between IS6100 and linA is the longest in UT26 (Fig. 7A), and the linB gene in UT26 has no IS6100 element in its flanking regions (Fig. 7B). Moreover, IS6100 is located at only one side of linC (Fig. 7C) and the linRED cluster (Fig. 7D) in UT26. These results suggested
Figure 6. (continued)
that UT26 is the closest to the prototype of the γ-HCH degrader, at least among the four strains examined in this study.

3.8. Conclusions and perspectives

Our comparison of the complete genome sequences of four γ-HCH-degrading sphingomonad strains and gathering of experimental data in this study demonstrate or strongly suggest the following points: (i) the gene repertoires and genomic organizations of the four γ-HCH-degrading strains, which are phylogenetically dispersed among related sphingomonad strains (Fig. 2), are relatively different from one another (Table 1 and Fig. 3); (ii) all four strains carry almost identical linA to linE genes for the conversion of γ-HCH to maleylacetate (Fig. 1 and Table 2); (iii) considerably different genes are used for the metabolism of maleylacetate in TKS (Fig. 1, Table 2, and
the ABC transporter necessary for γ-HCH utilization are structurally divergent, and such divergence reflects the phylogenetic relationship of their hosts (Fig. 2 and Supplementary Fig. S5 and Table 2); (v) most of the linA to linJ genes for the catabolic enzymes are located on several replicons whose replication/partition systems are highly conserved among sphingomonad plasmids (Tables 1 and 3); and (vi) the transposition of Is6100 has caused dynamic genome rearrangements including the fusion and resolution of replicons and the diversification of lin-flanking regions in the four strains (Figs. 6 and 7).

Based on our results in this study, we propose that these γ-HCH-degraders were formed independently in different geographic regions through the recruitment of specific lin genes and genes for the metabolism of maleylacetate into ancestral strains that had the core functions (including the linKLMN-encoded one) of sphingomonads. Multiple plasmids whose replication/partition machineries are highly conserved in sphingomonads might have played important roles in the recruitment of the specific lin genes by their horizontal transfer. In addition, Is6100 likely plays a crucial ‘editing’ role in the distribution and organization of the lin genes in genomes. In the future, our hypothesis may be confirmed in experiments using the four HCH degraders and their related but non-HCH-degrading and/or Is6100-free sphingomonad strains.

4. Data availability

The sequences with the annotation of replicons in MM-1, MI1205, and TKS have been deposited in DDBJ/EMBL/GenBank databases under the accession numbers shown in Table 1. Nucleotide sequences of the Tn3-type transposons, Tn6268 to Tn6278, were deposited in the DDBJ/EMBL/GenBank databases under accession numbers LC102249 to LC102259, respectively.

Conflict of interest

None declared.

Accession number

CP004036 to CP004041, CP005083 to CP005093, CP005188 to CP005193, and LC102249 to LC102259.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

References

1. Ogata, Y., Takada, H., Mizukawa, K., et al. 2009. International Pellet Watch: global monitoring of persistent organic pollutants (POPs) in coastal waters. 1. Initial phase data on PCBs, DDTs, and HCHs, Mar. Pollut. Bull., 58, 1437–46.

2. El-Shahawi, M. S., Hamza, A., Bashammakh, A. S. and Al-Saggaf, W. T. 2010, An overview on the accumulation, distribution, transformations, toxicity and analytical methods for the monitoring of persistent organic pollutants, Talanta, 80, 1587–97.

3. Tarcu, D., Cucu-Man, S., Borurkova, J., Klanova, J. and Covaci, A. 2013, Organochlorine pesticides in soil, moss and tree-bark from North-Eastern Romania, Sci. Total Environ., 456–457, 317–24.

4. Janssen, D. B., Dinkla, I. J., Poelarends, G. J. and Terpstra, P. 2003, Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities, Environ. Microbiol., 7, 1868–82.

5. Copley, S. D. 2009, Evolution of efficient pathways for degradation of anthropogenic chemicals, Nature Chem. Biol., 5, 559–66.

6. Stolz, A. 2009, Molecular characteristics of xenobiotic-degrading sphingomonads, Appl. Microbiol. Biotechnol., 81, 793–811

7. Lal, R., Pandey, G., Sharma, P., et al. 2010, Biochemistry of microbial degradation of hexachlorocyclohexane and prospects for bioremediation, Microbiol. Mol. Biol. Rev., 74, 58–80.

8. Stolz, A. 2014, Degradative plasmids from sphingomonads, FEMS Microbiol. Lett., 350, 9–19.

9. Nagata, Y., Tabata, M., Ohsubo, Y. and Tsuda, M. 2015, Biodegradation of organochlorine pesticides. In: M. Yates, C. Nakatsu, R. Miller, S. Pillai (eds), Manual of Environmental Microbiology, 4th edn., ASM Press, Washington, DC, pp. 5.1.2–1–30.

10. Top, E. M. and Springael, D. 2003, The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds, Curr. Opin. Biotechnol., 14, 262–9.

11. Copley, S. D., Rokicki, J., Turner, P., Daligault, H., Nolan, M. and Land, M. 2011, The whole genome sequence of Sphingobium chlorophenolicum L-1: insights into the evolution of the pentachlorophenol degradation pathway, Genome Biol. Evol., 4, 184–98.

12. Udikovic-Kolic, N., Scott, C. and Martin-Laurent, F. 2012, Evolution of atrazine-degrading capabilities in the environment, Appl. Microbiol. Biotechnol., 96, 1175–89.

13. Satola, B., Wannier, J. H. and Steinbacher, A. 2013, Metabolic characteristics of the species Variovorax paradoxus, Appl. Microbiol. Biotechnol., 97, 541–60.

14. Phillips, T. M., Seech, A. G., Lee, H. and Trevors, J. T. 2003, Biodegradation of hexachlorocyclohexane (HCH) by microorganisms, Biodegradation, 16, 363–92.

15. Vigen, J., Abulash, P. C., Li, Y. F., et al. 2011, Hexachlorocyclohexane (HCH) as new Stockholm Convention POPs—a global perspective on the management of Lindane and its waste isolomers, Environ. Sci. Pollut. Res. Int., 18, 152–62.

16. Nagata, Y., Endo, R., Ito, M., Ohsubo, Y. and Tsuda, M. 2007, Aerobic degradation of lindane (γ-hexachlorocyclohexane) in bacteria and its biochemical and molecular basis, Appl. Microbiol. Biotechnol., 76, 741–52.

17. Nagata, Y., Ohsubo, Y., Endo, R., et al. 2010, Complete genome sequence of the representative γ-hexachlorocyclohexane-degrading bacterium Sphingobium japonicum UT26, J. Bacteriol., 192, 5852–3.

18. Nagata, Y., Natsui, S., Endo, R., et al. 2011, Genomic organization and genomic structural rearrangements of Sphingobium japonicum UT26, an archetypal γ-hexachlorocyclohexane-degrading bacterium, Enzyme Microb. Technol., 49, 499–508.

19. Nagata, Y., Tabata, M., Ohhata, S., and Tsuda, M. 2014, Appearance and evolution of γ-hexachlorocyclohexane-degrading bacteria, In: H. Nojiri, M. Tsuda, M. Fukuda, Y. Kamagata (eds) Biodegradative Bacteria, Springer Verlag: Tokyo, pp. 19–41.

20. Verma, H., Kumar, R., Oldach, P., et al. 2014, Comparative genomic analysis of nine Sphingobium strains: insights into their evolution and hexachlorocyclohexane (HCH) degradation pathways, BMC genomics, 15, 1014.

21. Pearce, S. L., Oakeshott, J. G. and Pandey, G. 2015, Insights into ongoing evolution of the hexachlorocyclohexane catabolic pathway from comparative genomics of ten Sphingomonadaceae strains, G3, 5, 1081–94.
22. Tabata, M., Endo, R., Ito, M., et al. 2011, The lin genes for 3-hexachlorocyclohexane degradation in Sphingomonas sp. MM-1 proved to be dispersed across multiple plasmids, Biologia. Bacteriol. Biochem., 75, 466–72.

23. Tabata, M., Ohtsubo, Y., Ohhata, S., Tsuda, M. and Nagata, Y. 2013, Complete genome sequence of the γ-hexachlorocyclohexane-degrading bacterium Sphingobium sp. strain MM-1, Genome Announc., 1, e00247–13.

24. Ito, M., Prokop, Z., Klvana, M., et al. 2007, Degradation of β-hexachlorocyclohexane by haloalkane dehalogenase LinB from γ-hexachlorocyclohexane-utilizing bacterium Sphingobium sp. MI1205, Arch. Microbiol., 188, 313–25.

25. Tabata, M., Ohhata, S., Nikawadori, Y., et al. 2016, Complete genome sequence of a γ-hexachlorocyclohexane-degrading bacterium, Sphingobium sp. strain MI1205, Genome Announc., 4, e00246–16.

26. Tabata, M., Ohhata, S., Kawasumi, T. et al. 2016, Complete genome sequence of a γ-hexachlorocyclohexane degrader, Sphingobium sp. strain TKS, isolated from a γ-hexachlorocyclohexane-degrading microbial community, Genome Announc., 4, e00247–16.

27. Yabuuchi, E., Kosako, Y., Fujiwara, N., et al. 2002, Emendation of the genus Sphingobium Yabuuchi et al. 1990 and junior objective synonymy of the species of three genera, Sphingobium, Novosphingobium and Sphingopyxis, in conjunction with Blastomonas urincola, Int. J. Syst. Evol. Microbiol., 52, 1485–96.

28. Maniatis, T., Fritsch, E. and Sambrook, J. 1982, Molecular cloning: a Laboratory Manual. Cold Spring Harbor Laboratory: NY.

29. Endo, R., Ohtsubo, Y., Tsuda, M. and Nagata, Y. 2007, Identification and characterization of genes encoding a putative ABC-type transporter essential for utilization of γ-hexachlorocyclohexane in Sphingobium japonicum UT26, J. Bacteriol., 189, 3712–20.

30. Nagata, Y., Futamura, A., Miyachi, K. and Takagi, M. 1999, Two different types of dehalogenases, LinA and LinB, involved in γ-hexachlorocyclohexane degradation in Sphingomonas paucimobilis UT26 are localized in the periplasmic space without molecular processing, J. Bacteriol., 181, 5409–13.

31. Kovach, M. E., Elzer, P. H., Hill, D. S., et al. 1995, Four new derivatives of the broad-host-range cloning vector pBR1MCS, carrying different antibiotic-resistance cassettes, Gene, 166, 175–3.

32. Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J., et al. 1998, A plasmid pWDL7::cct in chloroaniline catabolism as determined by genomic and functional analyses, J. Bacteriol., 180, 4972–80.

33. Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y. and Tsuda, M. 2008, Distribution of -hexachlorocyclohexane degrader of hexachlorocyclohexane in Sphingomonas sp. strain KKS102, a versatile polychlorinated biphenyl/biphenyl degradation, J. Bacteriol., 190, 653–64.

34. Endo, R., Ohtsubo, Y., Tsuda, M. and Nagata, Y. 2007, Identification and characterization of genes for 2,4-dichlorophenoxyacetic acid degradation, plasmid pNB8c in chloroaniline catabolism as determined by genomic and functional analyses, Microb. Environ., 22, 348–56.

35. Krol, J. E., Penrod, J. T., McCaslin, H., et al. 2012, Role of IncP-1beta plasmid pEST4011 contains a novel IncP1 backbone and a catabolic ring-cleavage pathways and related 3-hydroxybenzoate and 4-hydroxybenzoate peripheral pathways in Burkholderia xenovorans LB400, Proc. Natl. Acad. Sci. USA, 109, 15280–7.

36. Lopez-Silva, M. J., Mendez, V., Agullo, L. and Segger, M. 2013, Genomic and functional analysis of the genovir and protocatechuate ring-cleavage pathways and related 3-hydroxybenzoate and 4-hydroxybenzoate peripheral pathways in Burkholderia xenovorans LB400, PLoS One, 8, e56038.

37. Stanier, R. Y., Palleroni, N. J. and Doudoroff, M. 1966, The aerobic pseudomonads: a taxonomic study, J. Gen. Microbiol., 43, 159–271.

38. Yohara, S., Komatsu, H., Goto, H., Ohtsubo, Y., Nagata, Y. and Tsuda, M. 2012, Pleiotropic roles of iron-responsive transcriptional regulator Fur in Burkholderia multivorans, Microbiology, 154, 1763–74.

39. Nagata, Y., Senbongi, J., Ishibashi, Y. et al. 2014, Identification of Burkholderia multivorans ATCC 17616 genetic determinants for fitness in soil by using signature-tagged mutagenesis, Microbiolology, 160, 883–91.

40. Perez-Pantoja, D., Domoso, R., Agullo, L., et al. 2012, Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales, Environ. Microbiol., 14, 1091–117.

41. Nelson, K. E., Wein, K., Paulsen, I. T., et al. 2002, Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440, Environ. Microbiol., 4, 799–808.

42. Miyachi, K., Lee, H. S., Fukuda, M., Takagi, M. and Nagata, Y. 2002, Cloning and characterization of Imk, involved in regulation of the downstream degradation pathway for γ-hexachlorocyclohexane in Sphingomonas paucimobilis UT26, Appl. Environ. Microbiol., 68, 1803–7.

43. Harwood, C. S. and Parales, R. E. 1996, The β-ketoisopropate pathway and the biology of self-identity, Annu. Rev. Microbiol., 50, 533–590.

44. Gross, R., Guzman, C. A., Sebahia, M., et al. 2008, The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetella, BMC genomics, 9, 449.

45. Vedler, E., Vater, M. and Hemaru, A. 2004, The completely sequenced plasmid pEST401 contains a novel IncP1 backbone and a catabolic transposon harboring tfd genes for 2,4-dichlorophenoxyacetic acid degradation, J. Bacteriol., 186, 7161–74.

46. Krol, J. E., Penrod, J. T., McCaslin, H., et al. 2012, Role of IncP-1beta plasmid pWDL7::ctf and pNB8c in chloroaniline catabolism as determined by genomic and functional analyses, Appl. Environ. Microbiol., 78, 828–38.

47. Mukhterje, U., Kumar, R., Mahato, N. K., Khurana, J. P. and Lal, R. 2013, Draft genome sequence of Sphingobium sp. strain HDIPO4, an avid degrader of hexachlorocyclohexane, Genome Announc., 1, e00749–13.

48. Davison, J. 1999, Genetic exchange between bacteria in the environment, Plasmid, 42, 73–91.

49. Shintani, M. and Nojiri, H. 2013, Mobile genetic elements (MGEs) carrying catabolic genes, In: A., Malik, E., Grohmann, M., Alves (eds), Management of Microbial Resources in the Environment, Springer: the Netherlands, pp. 167–214.

50. Ohtsubo, Y., Ishibashi, Y., Naganawa, H., et al. 2012, Conjugal transfer of plasmid originated bifunctional degradation genes in Acidovorax sp. strain KKS102, which are located on an integrative and conjugative element, J. Bacteriol., 194, 4237–48.

51. Shirvastava, N., Prokop, Z. and Kumar, A. 2015, Novel LinA-type 3-3-hexachlorocyclohexane dehydrochlorinase, Appl. Environ. Microbiol., 81, 7535–9.
62. Moriuchi, R., Tanaka, H., Nikawadori, Y., et al. 2014, Stepwise enhancement of catalytic performance of haloalkane dehalogenase LinB towards β-hexachlorocyclohexane, *AMB Express*, 4, 72.

63. Pandey, R., Lucent, D., Kumari, K., et al. 2014, Kinetic and sequence-structure-function analysis of LinB enzyme variants with β- and δ-hexachlorocyclohexane, *PLoS One*, 9, e103632.

64. Nagata, Y., Ohtsubo, Y. and Tsuda, M. 2015, Properties and biotechnological applications of natural and engineered haloalkane dehalogenases, *Appl. Microbiol. Biotechnol.*, 99, 9865–81.

65. Brassinga, A. K. and Marczynski, G. T. 2001, Replication intermediate analysis confirms that chromosomal replication origin initiates from an unusual intergenic region in *Caulobacter crescentus*, *Nucleic Acids Res.*, 29, 4441–51.

66. Sibley, C. D., MacLellan, S. R. and Finan, T. 2006, The *Sinorhizobium meliloti* chromosomal origin of replication, *Microbiology*, 152, 443–55.

67. Schaper, S. and Messer, W. 1995, Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target, *J. Biol. Chem.*, 270, 17622–6.

68. Pinto, U. M., Pappas, K. M. and Winans, S. C. 2012, The ABCs of plasmid replication and segregation, *Nat. Rev. Microbiol.*, 10, 755–65.

69. Shintani, M., Urata, M., Inoue, K., et al. 2007, The *Sphingomonas* plasmid pCAR3 is involved in complete mineralization of carbazole, *J. Bacteriol.*, 189, 2007–20.

70. Miller, T. R., Delcher, A. L., Salzberg, S. L., Saunders, E., Dettter, J. C. and Halden, R. U. 2010, Genome sequence of the dioxin-mineralizing bacterium *Sphingomonas wittichii* RW1, *J. Bacteriol.*, 192, 6101–2.

71. Luo, Y. R., Kang, S. G., Kim, S. J., et al. 2012, Genome sequence of benzo(a)pyrene-degrading bacterium *Novosphingobium pentaromativorans* US6-1, *J. Bacteriol.*, 194, 907.