Lessons from the genomes of lindane-degrading sphingomonads

Yuji Nagata,* Hiromi Kato, Yoshiyuki Ohtsubo and Masataka Tsuda

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

* For correspondence. E-mail aynaga@ige.tohoku.ac.jp Tel. (+81) 22 217 5682; Fax (+81) 22 217 5682.

Keywords Biodegradation, Genome, Mobile genetic element, Plasmid, Insertion sequence, Organochlorine pesticide, Xenobiotic, Sphingomonads

Summary

Bacterial strains capable of degrading man-made xenobiotic compounds are good materials to study bacterial evolution toward new metabolic functions. Lindane (γ-hexachlorocyclohexane, γ-HCH, or γ-BHC) is an especially good target compound for the purpose, since it is relatively recalcitrant but can be degraded by a limited range of bacterial strains. A comparison of the complete genome sequences of lindane-degrading sphingomonad strains clearly demonstrated that: (i) lindane-degrading strains emerged from a number of different ancestral hosts that have recruited *lin* genes encoding enzymes that are able to channel lindane to central metabolites, (ii) in sphingomonads *lin* genes have been acquired by horizontal gene transfer mediated by different plasmids and in which IS*6100* plays a role in recruitment and distribution of genes, and (iii) IS*6100* plays a role in dynamic genome rearrangements providing genetic diversity to different strains and ability to evolve to other states. Lindane-degrading bacteria whose genomes change so easily and quickly are also fascinating starting materials for tracing the bacterial evolution process experimentally in a relatively short time period. As the origin of the specific *lin* genes remains a mystery, such genes will be useful probes for exploring the cryptic "gene pool" available to bacteria.

Introduction

Every day, numerous chemical compounds are released into the environment by human activities. This dissemination often has serious environmental consequences, since most of these chemicals are not readily degraded in the environment and have harmful effects on humans and the natural ecosystem (Ogata *et al.*, 2009; El-Shahawi *et al.*, 2010; Tarcau *et al.*, 2013). Bacteria that degrade environmental pollutants have been isolated and characterized for the bioremediation of these toxic compounds, and have also attracted attention for their potential to be evolutionarily adapted to degrade chemical compounds unfamiliar to them (Janssen *et al.*, 2005; Copley, 2009; Stolz, 2009; Nagata *et al.*, 2016). Today, with increasingly large numbers of bacterial genomes and metagenomes being sequenced, it has become possible to discuss the evolution process of such bacteria, which in many ways remains shrouded in mystery.

The degradability of environmental pollutant varies widely from substance to substance. Simple aromatic compounds, *e.g.*, benzene, toluene, phenol, and naphthalene, are major environmental pollutants and are relatively easily degraded by microorganisms (Janssen *et al.*, 2005; Fuchs *et al.*, 2011; Perez-Pantoja *et al.*, 2012; Diaz *et al.*, 2013; Abbasian et al., 2016). In fact, many bacterial strains degrading such aromatic compounds have been isolated and studied in detail. In most cases, a series of genes encoding enzymes necessary for transformation of these compounds into TCA cycle intermediates constitute a gene cluster whose expression is often transcriptionally regulated (van der Meer *et al.*, 1992; Tropel and van der Meer, 2004; Diaz *et al.*, 2013; Kumar et al., 2016). In many cases genes form operons which are often located on mobile genetic elements, *e.g.*, transposons, plasmids, and integrative and conjugative elements (ICEs), and can be transferred between bacterial cells as a set (Top and Springael, 2003; van der Meer *et al.*, 1992; Tsuda *et al.*,

1999; Springael and Top, 2004; Liang *et al*., 2012; Ohtsubo *et al*., 2012); thus non-degrading bacterial cells can easily turn into degraders of aromatic compounds simply by acquiring such “ready-made” gene clusters. In other words, a system for degrading simple aromatic compounds has already been well established in nature, and the gene clusters necessary for the degradation can be distributed among bacterial cells in environments contaminated with these compounds, where cells having the ability to assimilate the compounds have a survival advantage. It is not surprising that the system for degrading simple aromatic compounds has been well established in nature, since most such compounds are not man-made but natural products and have existed for a long time in the environment.

On the other hand, anthropogenic compounds that were chemically synthesized or industrially produced are usually highly recalcitrant, because microorganisms have never or rarely encountered such chemical compounds and have not fully established systems to degrade and utilize them. However, bacteria that can degrade anthropogenic chemicals have been isolated, and most aerobic xenobiotics-degrading bacteria can use such chemicals as their sole sources of carbon and energy (Janssen *et al.*, 2005; Copley, 2009; Stolz, 2009; Nagata *et al.*, 2016; Hegedus *et al*., 2017; Nielsen *et al*.,2017; Singh, 2017). Since it has been proposed that the pathways for aerobic degradation of man-made xenobiotic compounds evolved relatively quickly within several decades after the release of such compounds into the environment, the bacterial strains capable of degrading man-made xenobiotic compounds are excellent models for studying the “primitive” adaptation and evolution processes of bacteria in the environment (Janssen *et al.*, 2005; Copley, 2009; Stolz, 2009; Nagata *et al.*, 2016).

Recent genome analyses of such xenobiotics-degrading strains have strongly suggested that they indeed emerged relatively recently by gathering genes for the degradation of xenobiotic compounds, and that mobile genetic elements played important roles for recruitment of the genes (Udikovic-Kolic *et al.*, 2012; Satola *et al.*, 2013; Nagata *et al.*, 2016). In this mini review, we will explain in detail our hypothesis for the emergence and evolution of the lindane (γ-hexachlorocyclohexane, γ-HCH, or γ-BHC) degraders by using the complete genome sequences of lindane-degrading sphingomonad strains, since the lindane degradation system in aerobic bacteria is an excellent model for investigating fundamental issues in microbial evolution (Nagata *et al.*, 2016; Tabata *et al.*, 2016c). Although the lindane-degrading bacterial strains belonging to non-sphingomonad groups have been reported (Lal *et al.*, 2010; Sineli *et al*., 2018), genes and enzymes for lindane degradation in non-sphingomonad bacteria remain unclear and their complete genome sequences are unavailable. Thus, we will focus on lindane-degrading strains belonging to sphingomonads in this article.

Sphingomonads, a bacterial group containing various strains degrading highly recalcitrant compounds

Many xenobiotics-degrading bacterial strains belonging to various taxonomic classifications have been isolated (Udikovic-Kolic *et al*., 2012; Satola *et al*., 2013), and "sphingomonads" are one of the most important bacterial groups for the degradation of recalcitrant hydrophobic compounds among bacteria that are widely distributed in the environment and can be easily cultured under laboratory conditions (Lal *et al.*, 2006; Stolz, 2009; Lal *et al.*, 2010; Stolz, 2014). Sphingomonads are a collective category comprising *Sphingomonas*, *Sphigobium*, *Novosphingobium*, and *Sphigopyxis* belonging to

105 *Alphaproteobacteria* (Yabuuchi *et al.*, 2002). Several sphingomonad strains have been
106 isolated that degrade highly recalcitrant hydrophobic compounds, *e.g.*, lindane (Tabata *et*
107 *al.*, 2016c; Verma *et al.*, 2017), pentachlorophenol (Copley *et al.*, 2011), dioxin-related
108 compounds (Miller *et al.*, 2010), lignin-related compounds (Masai *et al.*, 2007),
109 polyaromatic hydrocarbons (D'Argenio *et al.*, 2011), polyvinyl alcohol (Ohtsubo *et al.*,
110 2015a), polyethylene/polypropylene glycol (Ohtsubo *et al.*, 2015b; Ohtsubo *et al.*, 2015c;
111 Ohtsubo *et al.*, 2016a; Ohtsubo *et al.*, 2016b), and organophosphate (Parthasarathy *et al.*,
112 2017).

113 It is suggested that sphingomonads can adapt quickly or efficiently to the degradation
114 of new compounds in the environment, and it should be noted that each sphingomonad
115 strain degrading a highly recalcitrant compound cannot degrade any other highly
116 recalcitrant compounds. For example, lindane degraders cannot degrade
117 pentachlorophenol, dioxin-related compounds, and so on. In addition, there are also
118 sphingomonad strains that degrade no special compound, and such strains are often found
119 in the environment (Lauro *et al.*, 2009). On the basis of these facts, it can be speculated
120 that most sphingomonad strains in the environment are “ordinary”, but have the potential
121 to become “specialists” for the degradation of highly recalcitrant hydrophobic compounds.

122 Primary comparison of the genome sequences of sphingomonad strains including
123 degraders of highly recalcitrant compounds supports the idea that sphingomonads are a
124 versatile group because of the plasticity of their genomes (Nagata *et al.*, 2011; Aylward *et*
125 *al.*, 2013). It is strongly suggested that plasmid-mediated gene transfer and chromosome-
126 plasmid recombination, together with prophage and transposon-mediated rearrangements,
127 play prominent roles in the genome evolution of sphingomonads (Copley *et al.*, 2011;
128 Nagata *et al.*, 2011; Tabata *et al.*, 2016; Hegedus *et al.*, 2017). In some cases, the gene

organizations seem to be edited by using insertion sequences. These points will be explained in greater detail in the following sections by using aerobic lindane-degrading sphingomonad strains.

Lindane-degrading sphingomonad strains

Lindane is a completely man-made chlorinated pesticide that has caused serious environmental problems due to its toxicity and long persistence in upland soils (Phillips *et al.*, 2005; Vijgen *et al.*, 2011; Lal *et al.*, 2010). Although the use of lindane is now banned in most countries, this compound still remains in various environments and causes serious environmental problems (Vijgen *et al.*, 2011). Lindane is chemically synthesized by the process of photochlorination of benzene. The synthesized product is called technical-HCH (t-HCH) and consists mainly of five isomers, α- (60-70%), γ- (12-16%), β- (10-12%), δ- (6-10%), and ε-HCH (3-4%) (Vijgen *et al.*, 2011). Among these isomers, only γ-HCH has insecticidal activity, and it is used after purification as the insecticide lindane (> 99% purity). The remaining isomers have often been improperly disposed of, causing serious environmental problems, and thus, in addition to γ-HCH, α- and β-HCH isomers were also included as persistent organic pollutants (POPs) that must be controlled under international agreement at the Stockholm Convention (Vijgen *et al.*, 2011). Among the HCH isomers, β-HCH is the most recalcitrant; it is usually the predominant isomer remaining in contaminated soils and in animal tissues and fluids (Willett *et al.*, 1998).

Only several decades after the first release of lindane into the environment, a number of bacterial strains that aerobically degrade lindane have been isolated from geographically dispersed locations, and most such strains—particularly those that have been intensively analyzed—are sphingomonads as reviewed by Lal *et al.* (2006; 2010). The lindane

degradation pathway catalyzed by LinA, LinB, LinC, LinD, LinE, LinF, LinGH, and LinJ has been revealed as shown in Fig. 1. The *lin* genes for the conversion of lindane to β-ketoadipate (Fig. 1: *linA* to *linF*) are peculiar to the lindane-degrading pathway, since the β-ketoadipate pathway is often used by environmental bacterial strains for the assimilation of aromatic compounds (Harwood and Parales, 1996). It should be noted that the *linA* gene does not show significant similarity to any sequences in the databases except for the almost identical *linA* genes (> 90% identical) from lindane-degrading bacterial strains and metagenomes of HCH-polluted environments (Nagata *et al*., 2007; Lal *et al*., 2010).The *linA*, *linB*, and *linC* genes do not constitute an operon and are constitutively expressed at a relatively high level in UT26, while the *linD* and *linE* genes constitute an operon, and their expression is regulated by an LysR-type transcriptional regulator (LinR) (Miyauchi *et al*., 2002).

In addition to catabolic enzymes, a putative ABC-type transporter system consisting of four components (Fig. 1): permease, ATPase, periplasmic protein, and lipoprotein, encoded by *linK*, *linL*, *linM*, and *linN*, respectively, is necessary for the γ-HCH utilization in UT26 (Endo *et al*., 2007). The LinKLMN system is involved in γ-HCH utilization by conferring tolerance toward a toxic metabolite 2,5-dichlorophenol (Endo *et al*., 2006; Endo *et al*., 2007). The LinKLMN system is not a simple efflux pump of the toxic compound, but seems to be involved in the integrity of the outer membrane (Endo *et al*., 2007). It remains unknown how the LinKLMN system is involved in the integrity of the outer membrane, but the periplasmic protein LinM has a mammalian cell entry (Mce) domain (Casali and Riley, 2007), which is necessary for the lipid binding (Awai *et al*., 2006), and thus it is speculated that the LinKLMN system transports lipid-related compounds, *e.g.*, sphingolipid, for the integrity of the outer membrane.

Protein evolution of LinA and LinB

LinA and LinB are important targets from the viewpoint of protein evolution, since the LinA and LinB variants (> 90% identical) show different levels of enzymatic activity toward different HCH isomers and their metabolites (Nagata *et al*., 2007; Lal *et al*., 2010; Sharma, *et al*., 2011; Pandey *et al*., 2014; Nagata *et al*., 2015). It is quite noteworthy that the majority of the sequence variations in the *linA* and *linB* variants are non-synonymous substitutions, which strongly suggests that the *linA* and *linB* genes are still evolving at high speed under strong selection pressures (Nagata *et al*., 2007; Lal *et al*., 2010; Nagata *et al*., 2015).

Strain B90A has two different variants, LinA1$_{B90A}$ (LinA1 from strain B90A: the same expression will be used hereafter) and LinA2$_{B90A}$ (Kumari *et al*., 2002). LinA2$_{B90A}$ is identical to LinA$_{UT26}$, and LinA1$_{B90A}$ is 88% identical to LinA2$_{B90A}$/LinA$_{UT26}$. The differences between these two enzymes are caused mainly by the insertion of IS*6100* into the 3' end of the *linA1* gene (Kumari *et al*., 2002), and these two LinA variants are 93% identical in the N-terminal 148 amino acid region (11 amino-acid differences). In spite of this high similarity, LinA1$_{B90A}$ preferentially converts the (+) enantiomer of α-HCH, whereas LinA2$_{B90A}$ prefers the (-) enantiomer (Suar *et al*., 2005). The crystal structure of another LinA variant (LinA-type 2) was solved (Macwan *et al*., 2012). The gene for LinA-type 2, which was isolated from the metagenomic analysis, has no IS*6100* insertion at the 3' end (Macwan *et al*., 2012). However, LinA-type 2 is almost identical to LinA1$_{B90A}$ in the 1-148 amino acid region (only one amino acid difference), and also prefers the (+) enantiomer of α-HCH (Suar *et al*., 2005). Detailed kinetic and sequence-structure-function analyses of the seven naturally occurring LinA variants towards α-, γ-, and δ-HCH clearly

showed the contribution of sequence-structure differences to the differences in their stereospecificities for these HCH isomers and enantiospecificities for (+)- and (-)-α-HCH (Sharma, *et al.*, 2011).

Functional analyses of LinB variants have also been done, especially with respect to β-HCH degradation activity (Okai *et al.*, 2013). As mentioned above, β-HCH is the most recalcitrant among the four major isomers of t-HCH (Willett *et al.*, 1998). Its six chlorines are all in equatorial positions, which seem to confer the greatest chemical stability to this isomer. $LinB_{UT26}$ converts β-HCH to 2,3,4,5,6-pentachlorocyclohexanol (PCHL) (Nagata *et al.*, 2005), but it cannot be further metabolized. On the other hand, $LinB_{MI1205}$ from *Sphingobium* sp. MI1205 [identical to $LinB_{B90A}$ from *Sphingobium indicum* B90A (Sharma *et al.*, 2006) and $LinB_{BHC-A}$ from *Sphingomonas* sp. BHC-A (Wu *et al.*, 2007)], which is 98% identical (having a difference in only 7 of the 296 amino acid residues) to $LinB_{UT26}$, can catalyze the two-step conversion of β-HCH to 2,3,5,6-tetrachlorocyclohexane-1,4-diol (TCDL) with the first conversion step being an order of magnitude more rapid than that by $LinB_{UT26}$ (Ito *et al.*, 2007). Analysis of intermediate mutants between $LinB_{UT26}$ and $LinB_{MI1205}$ demonstrated that the β-HCH degradation activity of $LinB_{UT26}$ can be enhanced in a stepwise manner by the accumulation of point mutations (Moriuchi *et al.*, 2014). In addition to $LinB_{MI1205}$ and $LinB_{UT26}$, several other natural variants are known (Nagata *et al.*, 2007; Lal *et al.*, 2010; Moriuchi *et al.*, 2014; Pandey *et al.*, 2014). A comprehensive analysis of naturally occurring and synthetic variants of LinB with specific degradative activity toward HCH isomers was performed (Pandey *et al.*, 2014). One of the synthetic mutants that was constructed based on the data for naturally occurring LinB variants showed nearly 80-fold higher activity toward β- and δ-HCH than $LinB_{MI1205}$, clearly indicating the activity of LinB can be further improved (Pandey *et al.*, 2014).

Genomes of lindane-degrading sphingomonad strains

The complete genome sequence of UT26 was first determined among lindane-degraders (Nagata *et al*., 2010; Nagata *et al*., 2011), and now the complete genome sequences of four other γ-HCH-degraders are available: *Sphigobium* sp. MI1205 from Miyagi, Japan (Ito *et al*., 2007; Tabata *et al*., 2016a), *Sphingomonas* sp. MM-1 from India (Tabata *et al*., 2011; Tabata *et al*., 2013), *Sphingobium* sp. TKS from Kyushu, Japan (Tabata *et al*., 2016b), and *Sphingobium indicum* B90A (Verma *et al*., 2017). The genome organizations of these five lindane-degrading strains are summarized in Table 1. Strain B90A is another archetypal lindane-degrading bacterium that has been deeply analyzed (Verma *et al*., 2017). This strain is phylogenetically very close to UT26 (Fig. 2), suggesting that these two strains may be derived from the same ancestral lindane degrader. On the other hand, UT26/B90A, MI1205, MM-1, and TKS appear to be phylogenetically dispersed on the basis of a 16S rRNA gene analysis among closely related sphingomonad strains (Fig. 2). Since UT26 and B90A are very similar at whole genome level (Verma *et al*., 2017), we mainly compared UT26 and other three strains. The gene repertoires of UT26, MI1205, MM-1, and TKS (each strain has 4,128 to 5,248 ORF clusters of the total 10,325 ORF clusters among the four strains) are quite different from each other (only 1,288 ORF clusters are shared) (Tabata et al., 2016c). These results clearly indicated that the four lindane degraders are phylogenetically divergent, and it was strongly suggested that each of them acquired its lindane degradation ability independently.

Lindane is degraded in MI1205, MM-1, TKS, and B90A by the same pathway as in UT26 (Tabata *et al*., 2016c; Verma *et al*., 2017; Fig. 1). All five strains carry almost identical *linA* to *linE* genes for the conversion of lindane to maleylacetate, and MI1205,

MM-1, and B90A also carry almost identical *linF* and *linGHIJ* genes (*linI* encodes IclR-family transcriptional regulator probably involved in the expression of *linGH* genes) for the metabolism of maleylacetate, while different genes that show no significant similarity to the *linF* and *linGHIJ* genes at the DNA level (*linFb* and *linGHIJ* homologues) are used for the latter conversion steps in TKS (Tabata *et al*., 2016c). The *linKLMN* genes for the putative ABC transporter necessary for lindane utilization exhibit sequence divergence at amino acid level, which reflects the phylogenetic relationship of their hosts. However, they seem to have the same function, since the *linKLMN* homologues of MM-1, which is phylogenetically the most distant strain from UT26, could complement the *linKLMN* function in UT26 (Tabata *et al*., 2016c). Moreover, the *linKLMN* homologues were found not only in lindane degraders but also in non-lindane-degrading sphingomonad strains (Nagata *et al*., 2011). These findings strongly suggest that the *linKLMN* system is one of the inherent functions necessary for lindane utilization in sphingomonads. In summary, it can be concluded that the *lin* genes for the utilization of lindane consist of three types of genes for (i) the "specific" pathway for lindane degradation, (ii) a common pathway for the degradation of chlorinated aromatic compounds (where more than one gene has been found for the function), and (iii) inherent function(s) in sphingomonads (Fig. 1).

The four strains also have putative genes for the degradation of aromatic compounds (Tabata *et al*., 2016c), but the numbers of such ORFs (62, 46, 27, and 25 for TKS, UT26, MI1205, and MM-1, respectively) are much smaller than those in the versatile recalcitrant pollutant degraders, *Cupriavidus necator* JMP134 (Perez-Pantoja *et al*., 2008; Lykidis *et al*., 2010) and *Burkholderia xenovorans* LB400 (Chain *et al*., 2006; Romero-Silva *et al*., 2013) (149 and 135 for JMP134 and LB400, respectively). In particular, those in three strains except TKS are even smaller than those in the typical metabolically versatile soil

bacterial strains *Burkholderia multivorans* ATCC 17616 (Stanier *et al*., 1966; Yuhara *et al*., 2008; Perez-Pantoja *et al*., 2012; Nagata *et al*., 2014) and *Pseudomonas putida* KT2440 (Nelson *et al*., 2002)(73 and 62 for KT2440 and ATCC 17616, respectively). No specific genes for the degradation of other highly recalcitrant compounds were found in their genomes. These results support our hypothesis that the lindane-degrading sphingomonad strains are "specialists" for lindane degradation.

Plasmids in sphingomonads

It is generally accepted that horizontal gene transfer (HGT) is an important mechanism of microbial adaptation and genomic evolution (van der Meer *et al*., 1992; Tsuda *et al*., 1999; Top and Springael, 2003; Springael and Top, 2004; Liang *et al*., 2012; Touchon *et al*.,; Millan, 2018; Partridge *et al*., 2018; Sun, 2018; Cheng *et al*., 2019). HGT between bacteria in natural habitats is largely mediated by mobile genetic elements (MGEs), *e.g.*, self-transmissible plasmids, transposons, integrons, IS elements, ICEs, and bacteriophages. Among such known MGEs, plasmids are particularly important for the rapid adaptation of bacteria towards xenobiotics (Davison, 1999; Liang *et al*., 2012; Shintani and Nojiri, 2013; Stolz, 2014), and genes for the degradation of recalcitrant compounds are also often located on plasmids (Martinez *et al*., 2001; Trefault *et al*., 2004; Stolz, 2014).

Although all five lindane-degrading strains carry almost identical specific *lin* genes (*linA* to *linF*), they are dispersed on multiple replicons in the five strains (Table 1). In UT26 and B90A, some of the specific *lin* genes are located on chromosomes. On the other hand, all the specific *lin* genes are dispersed on multiple plasmids in various combinations in TKS, MI1205, and MM-1, although additional copies of *linB* and *linC* are also located on Chr1 in TKS (Table 1). The important point is that there are various replicon types of

297 such plasmids carrying the specific *lin* genes (Table 2: see below). In other words, no
298 plasmid has been found carrying a whole set of the specific *lin* genes. These observations
299 indicate that these strains acquired these genes by HGT, but not acquired a whole set of
300 responsible genes at once by the simple conjugative transfer of plasmids and/or ICEs as the
301 cases of aromatic compound-degrading strains (Davison, 1999; Ohtsubo *et al*., 2012;
302 Shintani and Nojiri, 2013).

303 We designated replicons having *rrn* operon(s) as "chromosomes". Indeed, all the
304 main chromosomes (Chr1s) of the five lindane degraders have *Alphaproteobacterial*-
305 chromosome-type replication origins (*oriC*s) (Brassinga and Marczynski, 2001; Sibley *et*
306 *al*., 2006). However, the Chr2s of UT26, TKS, and MI1205 have the plasmid-type
307 replication and active partition systems (Nagata *et al*., 2011; Tabata *et al*., 2016c). On the
308 basis of their replication/partition systems, it seems better to categorize them as plasmids.
309 Indeed, in B90A, pSRL2, which has a replication/partition region almost identical to that
310 of Chr2 of UT26, carries no *rrn* operon, and pSRL1, which has a replication/partition
311 region identical to that of pCHQ1 of UT26, carries an *rrn* operon (Table 1). Plasmids
312 including these plasmid-type chromosomes from sphingomonads can be classified based
313 on the similarities of their RepA (DNA replication initiator) proteins (Table 2). The RepA
314 proteins of plasmids in sphingomonads show a very low level of similarity to those of
315 well-studied plasmids (*e.g*., IncP-1, F, IincP-7, and IncP-9 plasmids).

316 The sizes and gene contents of the same type plasmids, even ones having the
317 identical *repA* gene, are highly divergent (Table 2), suggesting that the plasmids in
318 sphingomonads underwent dynamic rearrangements. It was clearly indicated that the
319 replicons having highly conserved replication/partition genes are distributed among
320 sphingomonad strains with frequent recombination events including replicon fusion

(Tabata *et al.*, 2016c). Interestingly, six pISP4-type plasmids carry identical *repA* and *parA* genes, and five of them also have other types of *repA* genes (Table 2), suggesting a prevalent fusion event of replicons in the pISP4-type plasmids. It is noteworthy that all six pISP4-type plasmids carrying identical *repA* and *parA* genes contain the *lin* genes (Table 2), suggesting that this type of plasmid plays an important role in dissemination of the *lin* genes.

Among the sphingomonad plasmids listed in Table 2, conjugal transferability of pCHQ1 and pLB1 has been experimentally confirmed (Nagata *et al.*, 2006; Miyazaki *et al.*, 2006). Originally, pLB1, which carries two copies of *linB*, was isolated from HCH-contaminated soil using the exogenous plasmid isolation technique (Miyazaki *et al.*, 2006). Moreover, metagenomic analysis also suggested the importance of the horizontal transfer of the specific *lin* genes by plasmids for HCH degradation in the environment (Sangwan *et al.*, 2012). These facts strongly suggest that conjugative plasmids play important roles in the distribution of the specific *lin* genes under environmental conditions. Since the conjugation host range properties of pCHQ1 and pLB1 seem to be narrow (Nagata *et al.*, 2006; Miyazaki *et al.*, 2006), these conjugative plasmids may only contribute HGT among sphingomonads-related bacteria.

Genome "editing" role of IS*6100* in lindane degraders

IS*6100* is the most abundant in the UT26, B90A, TKS, MI1205, and MM-1 genomes (13, 26, 29, 24, and 15 copies, respectively: Table 1) among putative transposable elements, including insertion sequence (IS) elements and Tn*3*-type transposons, suggesting that IS*6100* can transpose and increase its copy number in these lindane degraders.

Transposition of IS*6100* was indeed detected by the IS entrapment experiments in UT26, TKS, MI1205, and MM-1 (Tabata *et al.*, 2016c).

IS*6100* (i) belongs to the IS*6* family, (ii) is 880 bp long and carries a transposase gene and 14-bp terminal inverted repeats (IR) at both ends, (iii) has no apparent preference of target specificity, and (iv) is a "replicative" IS element and causes its duplication with an 8-bp duplication of the target sequence by its transposition (Mahillon and Chandler, 1998). Transposition of IS*6100* can generate three types of DNA rearrangements: intra-molecular transposition with a deletion/resolution (intra-replicon 1) or inversion (intra-replicon 2) event, and inter-molecular transposition with a fusion/integration (inter-replicon) event (Fig. 3).

It is possible to infer the most plausible past events caused by transposition of IS*6100* by comparison of the regions just upstream and downstream of copies of IS*6100* on the basis of the IS*6100* transposition mechanism (Fig. 3). Not only simple transposition with inversion but also transposition accompanied with the fusion and resolution of replicons must have occurred by transposition of IS*6100* in the lindane degraders as schematically shown in Fig. 4. In addition to the transposition, homologous recombination between two copies of IS*6100* seemed to occur (Fig. 3), strongly suggesting that IS*6100* can contribute to the dynamic genome rearrangements in the lindane degraders (Tabata *et al.*, 2016c).

TKS was isolated from γ-HCH-enriched liquid cultivation of a microbial community from a sediment sample contaminated with HCH isomers (Tabata *et al.*, 2016b). Recently we found the previously inferred structures without IS*6100* in metagenome sequence of the enrichment culture from which TKS was isolated (unpublished data), suggesting that such IS*6100*-transposition events indeed occurred during the enrichment culture in liquid

368 medium and the repeated single-colony isolation processes on the solid medium. This fact
369 indicates that rapid genome evolution is occurring in bacteria and suggests that the genome
370 structure of the bacterial strain isolated in the laboratory may be different from the
371 ancestral strain inhabiting the environment.

372 IS*6100* is highly associated with *lin* genes (Boltner *et al*., 2005; Lal *et al*., 2006;
373 Fuchu *et al*., 2008; Lal *et al*., 2010). A plasmid pLB1 that carries an IS*6100*-composite
374 transposon containing two copies of *linB* was isolated by the exogenous plasmid isolation
375 technique (Miyazaki *et al.*, 2006). These facts suggest that IS*6100* plays an important role
376 in recruitment of the specific *lin* genes. Comparison of the specific *lin*-flanking regions in
377 the lindane-degrading strains revealed that not only the *lin* genes themselves but also their
378 flanking regions are highly conserved (Tabata *et al.*, 2016c). Interestingly, such conserved
379 regions are located very close to IS*6100*, and the distances between the IS*6100* copies and
380 the *lin* genes vary, indicating that IS*6100* is likely to play a crucial editing role in trimming
381 the regions unnecessary for lindane utilization and gathering the specific *lin* genes (Tabata
382 *et al*., 2016c: Fig. 4). This observation supports the “selfish operon model”, in which HGT
383 allows genes to cluster into an operon by a series of approximations (Lawrence and Roth,
384 1996; Lawrence, 1999). At least, the most plausible explanation is that the transposition of
385 IS*6100* led to the diversification of the distribution and organization of the *lin* genes in the
386 genomes.

387 The distance between IS*6100* and *linA* is the longest in UT26, and the *linB* gene in
388 UT26 has no IS*6100* element in its flanking regions. Moreover, IS*6100* is located at only
389 one side of *linC* and the *linRED* cluster in UT26 (Nagata *et al*., 2011; Tabata *et al*., 2016c).
390 These results suggested that UT26 is the closest to the prototype of the lindane degrader, at
391 least among the five strains whose complete genomes were determined (Table 1). In

addition, IS*6100* seems to be involved in the genetic instability of the specific *lin* genes. The *linA*, *linC*, and *linRED* genes in UT26 are genetically unstable, *i.e.*, spontaneous deletion mutants of the regions containing these genes could easily be obtained, and these deletion processes in the mutants can be most simply explained by the involvement of IS*6100* (Nagata *et al.*, 2011).

As in the case of IS*6100*, IS*1071*, a member of the Tn*3* family, is also often associated with the genes for the degradation of xenobiotics, including atrazine (Udikovic-Kolic *et al.*, 2012), 2,4-D (Liang *et al.*, 2012), and linuron (Dunon et al., 2018), suggesting that IS*1071* also has functions like those of IS*6100*. Generally, diverse IS family transposase genes are associated with genes for the degradation of xenobiotics (Liang *et al.*, 2012). It will be of great interest to learn how such combinations between IS elements and degradative genes have arisen.

Emergence and evolution of lindane-degrading bacteria

Comparison of the genomes of lindane-degrading sphingomonad strains strongly suggested that the lindane-degrading bacteria emerged through recruitment of the specific *lin* genes into an ancestral strain that had core functions of sphingomonads, which are inherent ones of this bacterial group and necessary for the assimilation of lindane, such as the LinKLMN-type ABC transporter system and the β-ketoadipate pathway (Fig. 5). Other unknown core functions may exist. One of the most important conclusions at present is that lindane-degraders seemed to emerge independently and in parallel around the world. 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 HGT. Along with the HGT, IS*6100* might have contributed to integration of

the specific *lin* genes into replicons that already existed in the ancestral sphingomonad strains. It is also speculated that IS*6100* is involved in the recruitment of the *lin* genes from an environmental "gene pool" whose details are still obscure.

Primitive lindane degraders seemed to be diversified by the involvement of IS*6100*, *e.g.*, through its transposition and homologous recombination between two copies of it, and other mutations (Fig. 5). As a result, the distribution and organization of the *lin* genes in genomes were diversified. Since most of the genomic regions of B90A and UT26 are highly conserved, B90A may be a strain emerged from a common ancestral lindane degrader with UT26 (Verma *et al*., 2017). However, the replicon organizations of these two strains are different from each other (Table 2), and the difference cannot be explained only by the involvement of IS*6100*. At least acquisition and/or loss of some plasmids seemed to have occurred during the diversification process of the two strains.

After continued diversification, selective pressure may produce the "evolved" lindane degraders in the future (Fig. 5). The *lin* system still seems to be evolving toward one or more optimal states, *e.g*., by gathering the *lin* genes into a single replicon, by forming an operon of the *lin* genes, and by the continued evolution of Lin enzymes. It is noteworthy that pMI1 is a replicon that has almost all the specific *lin* genes (Table 1), and if the *linA* gene is introduced into this plasmid, a replicon carrying all the genes encoding enzymes necessary for the conversion of lindane to TCA cycle intermediates will be created. On the other hand, it is also important that IS*6100* is involved in the loss of the *lin* genes as described above (Nagata *et al.*, 2011), and thus can contribute to the adaptation for other conditions under which the *lin* genes are no longer necessary.

Concluding remarks

The genome sequences of a vast number of bacterial strains have been determined, and it has become possible to discuss the emergence and evolution of bacterial strains that degrade xenobiotics on the basis of their genomic information. Here, we presented a hypothesis to explain how the lindane-degrading bacteria emerged and are evolving. In the future, our hypothesis may be confirmed in experiments using the lindane degraders and their related but non-lindane-degrading and/or IS*6100*-free sphingomonad strains. The genomes of lindane-degrading bacteria are also good examples of how readily and quickly the bacterial genomes are changing, suggesting that we are currently just observing "snap shots" of the bacterial genomes. It should be noted that the strains we isolated through enrichment culture and single colony isolation processes under laboratory conditions may be artificial strains that never existed in natural environments.

Currently many draft genome sequences of other HCH (including not only lindane but also other HCH isomers) degraders and their related but non-HCH-degrading strains are available. Their comparative analyses provided us some important primary information on the evolution of HCH-degraders with the involvement of plasmids and insertion sequences (Verma *et al*., 2014; Pearce *et al*., 2015). However, only analysis of the complete genome sequences can provide us some advanced information, *e.g*., (i) the genome organizations of lindane degraders, (ii) the localization of *lin* genes on their genomes, and (iii) how plasmids/insertion sequences are involved in the emergence and evolution of lindane degraders (Tabata *et al*., 2016c; Verma *et al*., 2017).

Lastly, the origin of the specific *lin* genes is still a mystery. Especially, the genetic origin of *linA* remains completely unknown, since no sequence has been found that shows significant similarity to the *linA* gene, despite the availability of a large number of nucleotide sequences including metagenomes. The *linA* gene was partially reconstructed *in*

vitro by using a technique called metagenomic DNA shuffling (Boubakri *et al*., 2006). In the reconstruction, 74% of the *linA* gene came from metagenomic DNA extracted from non-HCH-contaminated and *linA*-free soils. This interesting study demonstrated that even noncontaminated soils have the potential to create the *linA* gene. The *linA* gene might be created in the environment by bacterial adaptability to novel compounds through a DNA shuffling process. However, no evidence has been reported to date that a new gene was created via a combination of independent small DNA fragments in the environment, although it is generally accepted that genes for enzymes evolved via duplication and recombination of smaller functional elements (Peisajovich *et al*., 2006). Alternatively, we speculate that the *linA* gene already existed in the "gene pool" from which bacteria draw genes according to their need. The *linA* gene should be a useful probe for exploring this cryptic gene pool available to bacteria.

Acknowledgements This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Institute for Fermentation, Osaka (IFO), Japan. The authors declare no conflict of interest.

References

Abbasian, F., Lockington, R., Megharaj, M., and Naidu, R. (2016) A review on the genetics of aliphatic and aromatic hydrocarbon degradation. *Appl Biochem Biotechnol* **178**: 224-250.

Awai, K., Xu, C., Tamot, B., and Benning, C. (2006) A phosphatidic acid-binding protein of the chloroplast inner envelope membrane involved in lipid trafficking. *Proc Natl Acad Sci U S A* **103**: 10817-10822.

Aylward, F.O., McDonald, B.R., Adams, S.M., Valenzuela, A., Schmidt, R.A., Goodwin, L.A. et al. (2013) Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Appl Environ Microbiol* **79**: 3724-3733.

Boltner, D., Moreno-Morillas, S., and Ramos, J.L. (2005) 16S rDNA phylogeny and distribution of lin genes in novel hexachlorocyclohexane-degrading *Sphingomonas* strains. *Environ Microbiol* **7**: 1329-1338.

Boubakri, H., Beuf, M., Simonet, P., and Vogel, T.M. (2006) Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene* **375**: 87-94.

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**: 4441-4451.

Casali, N., and Riley, L.W. (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**: 60.

Chain, P.S., Denef, V.J., Konstantinidis, K.T., Vergez, L.M., Agullo, L., Reyes, V.L. et al. (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci U S A* **103**: 15280-15287.

Cheng, M., Li, Y., Ma, Y., Qiu, J., Yan, X., and He, J. (2018) Complete genome sequence of *Sphingobium baderi* DE-13, an alkyl-substituted aniline-mineralizing bacterium. *Curr*

Microbiol **75**: 27-31.

Copley, S.D. (2009) Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat Chem Biol* **5**: 559-566.

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**: 184-198.

D'Argenio, V., Petrillo, M., Cantiello, P., Naso, B., Cozzuto, L., Notomista, E. et al. (2011) *De novo* sequencing and assembly of the whole genome of *Novosphingobium* sp. strain PP1Y. *J Bacteriol* **193**: 4296.

Davison, J. (1999) Genetic exchange between bacteria in the environment. *Plasmid* **42**: 73-91.

Diaz, E., Jimenez, J.I., and Nogales, J. (2013) Aerobic degradation of aromatic compounds. *Curr Opin Biotechnol* **24**: 431-442.

Dunon, V., Bers, K., Lavigne, R., Top, E.M., and Springael, D. (2018) Targeted metagenomics demonstrates the ecological role of IS*1071* in bacterial community adaptation to pesticide degradation. *Environ Microbiol* **20**: 4091-4111.

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**: 1587-1597.

Endo, R., Ohtsubo, Y., Tsuda, M., and Nagata, Y. (2006) Growth inhibition by metabolites of γ-hexachlorocyclohexane in *Sphingobium japonicum* UT26. *Biosci Biotechnol Biochem* **70**: 1029-1032.

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**: 3712-3720.

529 Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds - from
530 one strategy to four. *Nat Rev Microbiol* **9**: 803-816.
531 Fuchu, G., Ohtsubo, Y., Ito, M., Miyazaki, R., Ono, A., Nagata, Y., and Tsuda, M. (2008)
532 Insertion sequence-based cassette PCR: cultivation-independent isolation of γ-
533 hexachlorocyclohexane-degrading genes from soil DNA. *Appl Microbiol Biotechnol* **79**: 627-
534 632.
535 Harwood, C.S., and Parales, R.E. (1996) The β-ketoadipate pathway and the biology of self-
536 identity. *Annu Rev Microbiol* **50**: 553-590.
537 Hegedus, B., Kos, P.B., Balint, B., Maroti, G., Gan, H.M., Perei, K., and Rakhely, G. (2017)
538 Complete genome sequence of *Novosphingobium resinovorum* SA1, a versatile xenobiotic-
539 degrading bacterium capable of utilizing sulfanilic acid. *J Biotechnol* **241**: 76-80.
540 Ito, M., Prokop, Z., Klvana, M., Otsubo, Y., Tsuda, M., Damborsky, J., and Nagata, Y. (2007)
541 Degradation of β-hexachlorocyclohexane by haloalkane dehalogenase LinB from γ-
542 hexachlorocyclohexane-utilizing bacterium *Sphingobium* sp. MI1205. *Arch Microbiol* **188**:
543 313-325.
544 Janssen, D.B., Dinkla, I.J., Poelarends, G.J., and Terpstra, P. (2005) Bacterial degradation of
545 xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ*
546 *Microbiol* **7**: 1868-1882.
547 Kumar, A., Trefault, N., and Olaniran, A.O. (2016) Microbial degradation of 2,4-
548 dichlorophenoxyacetic acid: Insight into the enzymes and catabolic genes involved, their
549 regulation and biotechnological implications. *Crit Rev Microbiol* **42**: 194-208.
550 Kumari, R., Subudhi, S., Suar, M., Dhingra, G., Raina, V., Dogra, C. et al. (2002) Cloning and
551 characterization of *lin* genes responsible for the degradation of hexachlorocyclohexane isomers
552 by *Sphingomonas paucimobilis* strain B90. *Appl Environ Microbiol* **68**: 6021-6028.

553 Lal, R., Dogra, C., Malhotra, S., Sharma, P., and Pal, R. (2006) Diversity, distribution and
554 divergence of *lin* genes in hexachlorocyclohexane-degrading sphingomonads. *Trends*
555 *Biotechnol* **24**: 121-130.
556 Lal, R., Pandey, G., Sharma, P., Kumari, K., Malhotra, S., Pandey, R. et al. (2010) Biochemistry
557 of microbial degradation of hexachlorocyclohexane and prospects for bioremediation.
558 *Microbiol Mol Biol Rev* **74**: 58-80.
559 Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S. et al. (2009) The
560 genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106**: 15527-
561 15533.
562 Lawrence, J. (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and
563 eukaryotes. *Curr Opin Genet Dev* **9**: 642-648.
564 Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution
565 of gene clusters. *Genetics* **143**: 1843-1860.
566 Liang, B., Jiang, J., Zhang, J., Zhao, Y., and Li, S. (2012) Horizontal transfer of dehalogenase
567 genes involved in the catalysis of chlorinated compounds: evidence and ecological role. *Crit*
568 *Rev Microbiol* **38**: 95-110.
569 Luo, Y.R., Kang, S.G., Kim, S.J., Kim, M.R., Li, N., Lee, J.H., and Kwon, K.K. (2012) Genome
570 sequence of benzo(a)pyrene-degrading bacterium *Novosphingobium pentaromativorans* US6-1.
571 *J Bacteriol* **194**: 907.
572 Lykidis, A., Perez-Pantoja, D., Ledger, T., Mavromatis, K., Anderson, I.J., Ivanova, N.N. et al.
573 (2010) The complete multipartite genome sequence of *Cupriavidus necator* JMP134, a versatile
574 pollutant degrader. *PLoS One* **5**: e9729.
575 Macwan, A.S., Kukshal, V., Srivastava, N., Javed, S., Kumar, A., and Ramachandran, R. (2012)
576 Crystal structure of the hexachlorocyclohexane dehydrochlorinase (LinA-type2): mutational

577 analysis, thermostability and enantioselectivity. *PLoS One* **7**: e50373.
578 Mahillon, J., and Chandler, M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62**: 725-774.
579 Martinez, B., Tomkins, J., Wackett, L.P., Wing, R., and Sadowsky, M.J. (2001) Complete
580 nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from
581 *Pseudomonas* sp. strain ADP. *J Bacteriol* **183**: 5684-5697.
582 Masai, E., Katayama, Y., and Fukuda, M. (2007) Genetic and biochemical investigations on
583 bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci Biotechnol*
584 *Biochem* **71**: 1-15.
585 Millan, A.S. (2018) Evolution of plamid-mediated antibiotic resistance in the clinical context. *Tre*
586 *Microbiol* **26**: 978-985.
587 Miller, T.R., Delcher, A.L., Salzberg, S.L., Saunders, E., Detter, J.C., and Halden, R.U. (2010)
588 Genome sequence of the dioxin-mineralizing bacterium *Sphingomonas wittichii* RW1. *J*
589 *Bacteriol* **192**: 6101-6102.
590 Miyauchi, K., Lee, H.S., Fukuda, M., Takagi, M., and Nagata, Y. (2002) Cloning and
591 characterization of *linR*, involved in regulation of the downstream pathway for γ-
592 hexachlorocyclohexane degradation in *Sphingomonas paucimobilis* UT26. *Appl Environ*
593 *Microbiol* **68**: 1803-1807.
594 Miyazaki, R., Sato, Y., Ito, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2006) Complete
595 nucleotide sequence of an exogenously isolated plasmid, pLB1, involved in γ-
596 hexachlorocyclohexane degradation. *Appl Environ Microbiol* **72**: 6923-6933.
597 Moriuchi, R., Tanaka, H., Nikawadori, Y., Ishitsuka, M., Ito, M., Ohtsubo, Y. et al. (2014)
598 Stepwise enhancement of catalytic performance of haloalkane dehalogenase LinB towards β-
599 hexachlorocyclohexane. *AMB Express* **4**: 72.
600 Nagata, Y., Endo, R., Ito, M., Ohtsubo, Y., and Tsuda, M. (2007) Aerobic degradation of lindane

(γ-hexachlorocyclohexane) in bacteria and its biochemical and molecular basis. *Appl Microbiol Biotechnol* **76**: 741-752.

Nagata, Y., Kamakura, M., Endo, R., Miyazaki, R., Ohtsubo, Y., and Tsuda, M. (2006) Distribution of γ-hexachlorocyclohexane-degrading genes on three replicons in *Sphingobium japonicum* UT26. *FEMS Microbiol Lett* **256**: 112-118.

Nagata, Y., Natsui, S., Endo, R., Ohtsubo, Y., Ichikawa, N., Ankai, A. 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.

Nagata, Y., Ohtsubo, Y., Endo, R., Ichikawa, N., Ankai, A., Oguchi, A. et al. (2010) Complete genome sequence of the representative γ-hexachlorocyclohexane-degrading bacterium *Sphingobium japonicum* UT26. *J Bacteriol* **192**: 5852-5853.

Nagata, Y., Ohtsubo, Y., and Tsuda, M. (2015) Properties and biotechnological applications of natural and engineered haloalkane dehalogenases. *Appl Microbiol Biotechnol* **99**: 9865-9881.

Nagata, Y., Prokop, Z., Sato, Y., Jerabek, P., Kumar, A., Ohtsubo, Y. et al. (2005) Degradation of β-hexachlorocyclohexane by haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26. *Appl Environ Microbiol* **71**: 2183-2185.

Nagata, Y., Senbongi, J., Ishibashi, Y., Sudo, R., Miyakoshi, M., Ohtsubo, Y., and Tsuda, M. (2014) Identification of *Burkholderia multivorans* ATCC 17616 genetic determinants for fitness in soil by using signature-tagged mutagenesis. *Microbiology* **160**: 883-891.

Nagata, Y., Tabata, M., Ohtsubo, Y., and Tsuda, M. (2016) Biodegradation of organochlorine pesticides. *Chapter 512 p 1-30 In Yates M, Nakatsu C, Miller R, Pillai S (ed), Manual of Environmental Microbiology, 4th Edition ASM Press, Washington, DC.*

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A. et al.

(2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799-808.

Nielsen, T.K., Rasmussen, M., Demaneche, S., Cecillon, S., Vogel, T.M., and Hansen, L.H. (2017) Evolution of sphingomonad gene clusters related to pesticide catabolism revealed by genome sequence and mobilomics of *Sphingobium herbicidovorans* MH. *Genome Biol Evol* **9**: 2477-2490.

Ogata, Y., Takada, H., Mizukawa, K., Hirai, H., Iwasa, S., Endo, S. 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**: 1437-1446.

Ohtsubo, Y., Ishibashi, Y., Naganawa, H., Hirokawa, S., Atobe, S., Nagata, Y., and Tsuda, M. (2012) Conjugal transfer of polychlorinated biphenyl/biphenyl degradation genes in *Acidovorax* sp. strain KKS102, which are located on an integrative and conjugative element. *J Bacteriol* **194**: 4237-4248.

Ohtsubo, Y., Nagata, Y., Numata, M., Tsuchikane, K., Hosoyama, A., Yamazoe, A. et al. (2015a) Complete genome sequence of polyvinyl alcohol-degrading strain *Sphingopyxis* sp. 113P3 (NBRC 111507). *Genome Announc* **3**.

Ohtsubo, Y., Nagata, Y., Numata, M., Tsuchikane, K., Hosoyama, A., Yamazoe, A. et al. (2015b) Complete genome sequence of *Sphingopyxis macrogoltabida* type strain NBRC 15033, originally isolated as a polyethylene glycol degrader. *Genome Announc* **3**.

Ohtsubo, Y., Nagata, Y., Numata, M., Tsuchikane, K., Hosoyama, A., Yamazoe, A. et al. (2015c) Complete genome sequence of polypropylene glycol- and polyethylene glycol-degrading *Sphingopyxis macrogoltabida* strain EY-1. *Genome Announc* **3**.

Ohtsubo, Y., Nonoyama, S., Nagata, Y., Numata, M., Tsuchikane, K., Hosoyama, A. et al. (2016a) Complete genome sequence of *Sphingopyxis terrae* strain 203-1 (NBRC 111660), a

649 polyethylene glycol degrader. *Genome Announc* **4**.
650 Ohtsubo, Y., Nonoyama, S., Nagata, Y., Numata, M., Tsuchikane, K., Hosoyama, A. et al. (2016b)
651 Complete genome sequence of *Sphingopyxis macrogoltabida* strain 203N (NBRC 111659), a
652 polyethylene glycol degrader. *Genome Announc* **4**.
653 Okai, M., Ohtsuka, J., Imai, L.F., Mase, T., Moriuchi, R., Tsuda, M. et al. (2013) Crystal structure
654 and site-directed mutagenesis analyses of haloalkane dehalogenase LinB from *Sphingobium* sp.
655 strain MI1205. *J Bacteriol* **195**: 2642-2651.
656 Pandey, R., Lucent, D., Kumari, K., Sharma, P., Lal, R., Oakeshott, J.G., and Pandey, G. (2014)
657 Kinetic and sequence-structure-function analysis of LinB enzyme variants with β- and δ-
658 hexachlorocyclohexane. *PLoS One* **9**: e103632.
659 Parthasarathy, S., Azam, S., Lakshman Sagar, A., Narasimha Rao, V., Gudla, R., Parapatla, H. et
660 al. (2017) Genome-guided insights reveal organophosphate-degrading *Brevundimonas diminuta*
661 as *Sphingopyxis wildii* and define its versatile metabolic capabilities and environmental
662 adaptations. *Genome Biol Evol* **9**: 77-81.
663 Partridge, S.R., Kwong, S.M., Firth, N., and Jensen, S.O. (2018) Mobile genetic elements
664 associated with antimicrobial resistance. *Clin Microbiol Rev* **31**: e00088-17.
665 Pearce, S.L., Oakeshott, J.G., and Pandey, G. (2015) Insights into ongoing evolution of the
666 hexachlorocyclohexane catabolic pathway from comparative genomics of ten
667 Sphingomonadaceae strains. *G3 (Bethesda)* **5**: 1081-1094.
668 Peisajovich, S.G., Rockah, L., and Tawfik, D.S. (2006) Evolution of new protein topologies
669 through multistep gene rearrangements. *Nat Genet* **38**: 168-174.
670 Perez-Pantoja, D., De la Iglesia, R., Pieper, D.H., and Gonzalez, B. (2008) Metabolic
671 reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-
672 degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol Rev* **32**: 736-794.

673 Perez-Pantoja, D., Donoso, R., Agullo, L., Cordova, M., Seeger, M., Pieper, D.H., and Gonzalez,
674 B. (2012) Genomic analysis of the potential for aromatic compounds biodegradation in
675 Burkholderiales. *Environ Microbiol* **14**: 1091-1117.
676 Phillips, T.M., Seech, A.G., Lee, H., and Trevors, J.T. (2005) Biodegradation of
677 hexachlorocyclohexane (HCH) by microorganisms. *Biodegradation* **16**: 363-392.
678 Romero-Silva, M.J., Mendez, V., Agullo, L., and Seeger, M. (2013) Genomic and functional
679 analyses of the gentisate and protocatechuate ring-cleavage pathways and related 3-
680 hydroxybenzoate and 4-hydroxybenzoate peripheral pathways in *Burkholderia xenovorans*
681 LB400. *PLoS One* **8**: e56038.
682 Sangwan, N., Lata, P., Dwivedi, V., Singh, A., Niharika, N., Kaur, J. et al. (2012) Comparative
683 Metagenomic analysis of soil microbial communities across three hexachlorocyclohexane
684 contamination Levels. *PLoS One* **7**: e46219.
685 Satola, B., Wubbeler, J.H., and Steinbuchel, A. (2013) Metabolic characteristics of the species
686 *Variovorax paradoxus*. *Appl Microbiol Biotechnol* **97**: 541-560.
687 Sharma, P., Pandey, R., Kumari, K., Pandey, G., Jackson, C.J., Russell, R.J. et al. (2011) Kinetic
688 and sequence-structure-function analysis of known LinA variants with different
689 hexachlorocyclohexane isomers. *PLoS One* **6**: e25128.
690 Sharma, P., Raina, V., Kumari, R., Malhotra, S., Dogra, C., Kumari, H. et al. (2006) Haloalkane
691 dehalogenase LinB is responsible for β- and δ-hexachlorocyclohexane transformation in
692 *Sphingobium indicum* B90A. *Appl Environ Microbiol* **72**: 5720-5727.
693 Shintani, M., and Nojiri, H. (2013) Mobile genetic elements (MGEs) carrying catabolic genes.
694 *Editors: Abdul Malik, Elisabeth Grohmann, Madalena Alves Book Title:Management of*
695 *Microbial Resources in the Environment. Springer, Netherlands*.
696 Sibley, C.D., MacLellan, S.R., and Finan, T. (2006) The *Sinorhizobium meliloti* chromosomal

origin of replication. *Microbiology* **152**: 443-455.

Singh, D. (2017) Biodegradation of xenobiotics-a way for environmental detoxification. *International Journal of Development Research* **7**: 14082-14087.

Sineli, P.E., Herrera, H.M., Cuozzo, S.A., and Davila Costa, J.S. (2018) Quantitative proteomic and transcriptional analyses reveal degradation pathway of γ-hexachlorocyclohexane and the metabolic context in the actinobacterium *Streptomyces* sp. M7. *Chemosphere* **211**: 1025-1034.

Springael, D., and Top, E.M. (2004) Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol* **12**: 53-58.

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

Stolz, A. (2009) Molecular characteristics of xenobiotic-degrading sphingomonads. *Appl Microbiol Biotechnol* **81**: 793-811.

Stolz, A. (2014) Degradative plasmids from sphingomonads. *FEMS Microbiol Lett* **350**: 9-19.

Suar, M., Hauser, A., Poiger, T., Buser, H.R., Muller, M.D., Dogra, C. et al. (2005) Enantioselective transformation of α-hexachlorocyclohexane by the dehydrochlorinases LinA1 and LinA2 from the soil bacterium *Sphingomonas paucimobilis* B90A. *Appl Environ Microbiol* **71**: 8514-8518.

Sun, D. (2018) Pull in and push out: mechanisms of horizontal gene transfer in bacteria. *Front Microbiol* **9**: 2154.

Tabata, M., Endo, R., Ito, M., Ohtsubo, Y., Kumar, A., Tsuda, M., and Nagata, Y. (2011) The *lin* genes for γ-hexachlorocyclohexane degradation in *Sphingomonas* sp. MM-1 proved to be dispersed across multiple plasmids. *Biosci Biotechnol Biochem* **75**: 466-472.

Tabata, M., Ohhata, S., Kawasumi, T., Nikawadori, Y., Kishida, K., Sato, T. et al. (2016b)

721 Complete genome sequence of a γ-hexachlorocyclohexane degrader, *Sphingobium* sp. strain
722 TKS, isolated from a γ-hexachlorocyclohexane-degrading microbial community. *Genome*
723 *Announc* **4**.
724 Tabata, M., Ohhata, S., Nikawadori, Y., Kishida, K., Sato, T., Kawasumi, T. et al. (2016c)
725 Comparison of the complete genome sequences of four γ-hexachlorocyclohexane-degrading
726 bacterial strains: insights into the evolution of bacteria able to degrade a recalcitrant man-made
727 pesticide. *DNA Res* **23**: 581-599.
728 Tabata, M., Ohhata, S., Nikawadori, Y., Sato, T., Kishida, K., Ohtsubo, Y. et al. (2016a) Complete
729 genome sequence of a γ-hexachlorocyclohexane-degrading bacterium, *Sphingobium* sp. strain
730 MI1205. *Genome Announc* **4**.
731 Tabata, M., Ohtsubo, Y., Ohhata, S., Tsuda, M., and Nagata, Y. (2013) Complete genome
732 sequence of the γ-hexachlorocyclohexane-degrading bacterium *Sphingomonas* sp. strain MM-1.
733 *Genome Announc* **1**.
734 Tarcau, D., Cucu-Man, S., Boruvkova, J., Klanova, J., and Covaci, A. (2013) Organochlorine
735 pesticides in soil, moss and tree-bark from North-Eastern Romania. *Sci Total Environ* **456-457**:
736 317-324.
737 Top, E.M., and Springael, D. (2003) The role of mobile genetic elements in bacterial adaptation to
738 xenobiotic organic compounds. *Curr Opin Biotechnol* **14**: 262-269.
739 Touchon, M., Moura de Sousa, J.A., and Rocha, E.P. (2017) Embracing the enemy: the
740 diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr*
741 *Opin Microbiol* **38**: 66-73.
742 Trefault, N., De la Iglesia, R., Molina, A.M., Manzano, M., Ledger, T., Perez-Pantoja, D. et al.
743 (2004) Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134
744 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of

745 specialized chloroaromatic degradation pathways. *Environ Microbiol* **6**: 655-668.
746 Tropel, D., and van der Meer, J.R. (2004) Bacterial transcriptional regulators for degradation
747 pathways of aromatic compounds. *Microbiol Mol Biol Rev* **68**: 474-500.
748 Tsuda, M., Tan, H.M., Nishi, A., and Furukawa, K. (1999) Mobile catabolic genes in bacteria. *J*
749 *Biosci Bioeng* **87**: 401-410.
750 Udikovic-Kolic, N., Scott, C., and Martin-Laurent, F. (2012) Evolution of atrazine-degrading
751 capabilities in the environment. *Appl Microbiol Biotechnol* **96**: 1175-1189.
752 van der Meer, J.R., de Vos, W.M., Harayama, S., and Zehnder, A.J. (1992) Molecular mechanisms
753 of genetic adaptation to xenobiotic compounds. *Microbiol Rev* **56**: 677-694.
754 Verma, H., Bajaj, A., Kumar, R., Kaur, J., Anand, S., Nayyar, N. et al. (2017) Genome
755 organization of *Sphingobium indicum* B90A: an archetypal hexachlorocyclohexane (HCH)
756 degrading genotype. *Genome Biol Evol* **9**: 2191-2197.
757 Verma, H., Kumar, R., Oldach, P., Sangwan, N., Khurana, J.P., Gilbert, J.A., and Lal, R. (2014)
758 Comparative genomic analysis of nine *Sphingobium* strains: insights into their evolution and
759 hexachlorocyclohexane (HCH) degradation pathways. *BMC Genomics* **15**: 1014.
760 Vijgen, J., Abhilash, P.C., Li, Y.F., Lal, R., Forter, M., Torres, J. et al. (2011)
761 Hexachlorocyclohexane (HCH) as new Stockholm Convention POPs-a global perspective on
762 the management of Lindane and its waste isomers. *Environ Sci Pollut Res Int* **18**: 152-162.
763 Willett, K., Ulrich, E., and Hites, R. (1998) Differential toxicity and environmental fates of
764 hexachlorocylohexane isomers. *Environ Sci Technol* **32**: 2197-2207.
765 Wu, J., Hong, Q., Han, P., He, J., and Li, S. (2007) A gene *linB2* responsible for the conversion of
766 β-HCH and 2,3,4,5,6-pentachlorocyclohexanol in *Sphingomonas* sp. BHC-A. *Appl Microbiol*
767 *Biotechnol* **73**: 1097-1105.
768 Yabuuchi, E., Kosako, Y., Fujiwara, N., Naka, T., Matsunaga, I., Ogura, H., and Kobayashi, K.

(2002) Emendation of the genus *Sphingomonas* Yabuuchi et al. 1990 and junior objective synonymy of the species of three genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, in conjunction with *Blastomonas ursincola*. *Int J Syst Evol Microbiol* **52**: 1485-1496.

Yuhara, S., Komatsu, H., Goto, H., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2008) Pleiotropic roles of iron-responsive transcriptional regulator Fur in *Burkholderia multivorans*. *Microbiology* **154**: 1763-1774.

Figure legends

Fig. 1. Functions necessary for lindane utilization in sphingomonads. Enzymes for the specific pathway for lindane degradation and for a common pathway for degradation of chlorinated aromatic compounds, *e.g.*, 2,4,5-T, 2,4-D, and PCP, are shown in red and blue, respectively. Genes for these functions are often located on plasmids and may be obtained by horizontal gene transfer. Note that β-ketoadipate pathway is common pathway for the degradation of aromatic compounds and is widely distributed among environmental bacteria. 1,2,4-TCB and 2,5-DCP are dead-end products, and 2,5-DCP has toxic effect on the cells, although these are experimentally confirmed only in UT26. The LinKLNM-type ABC transporter, which is an inherent function in sphingomonads, is involved in the tolerance for the toxic effect of 2,5-DCP. See text for detail.

Fig. 2. Phylogenetic tree of 16S rRNA genes of sphingomonad strains. Neighbor-joining phylogenetic tree of the conserved sites in 16S rRNA genes of 13 sphingomonad strains, *S. japonicum* UT26S (UT26_1, SJA_C1-r0010; UT26_2, SJA_C2-r0010; UT26_3, SJA_C2-r0040), *Sphingobium indicum* B90A (B90A_1, Chr_3240756_3242252; B90A_2, pSR1_46051_47546), *Sphingobium francense* Sp+ (Sp+, NR_042944), *Sphingobium* sp. TKS (TKS_1, Chr1_62351_63846; TKS_2, Chr2_117006_118503; and TKS_3, Chr2_376042_377539_c), *Sphingobium chlorophenolicum* L-1 (L-1_1, Sphch_R0043; L-2_2, Sphch_R0058; L-1_3, Sphch_R0067), *Sphingomonas* sp. SKA58 (SKA58_1, SKA58_r00366; SKA58_2, SKA58_r18278), *Sphigobium* sp. MI1205 (MI1205_1, Chr1_64638_66133; MI1205_2, Chr2_561355_562850_c), *Sphingobiums* sp. SYK-6 (SYK6_1, SLG_r0030; SYK6_2, SLG_r0060), *Sphingomonas wittichii* RW1 (RW1_1,

Swit_R0031; RW1_2, Swit_R0040), *Sphingomonas* sp. MM-1 (MM-1_1, Chr_1791835_1793331_c; MM-1_2, Chr_2084177_2085673_c), *Sphingopyxis alaskensis* RB2256 (RB2256, Sala_R0048), *Novosphingobium* sp. PP1Y (PPY_1, PP1Y_AR03; PPY_2, PP1Y_AR23 ;PPY_3, PP1Y_AR65), and *N. aromaticivorans* DSM 12444 (DSM_1, Saro_R0065; DSM_2, Saro_R0059; DSM_3, Saro_R0053) was constructed using MAFFT program (http://mafft.cbrc.jp/alignment/software/) and visualized by Njplot software. 16S rRNA gene (*rrsE*: gene ID 7437018) of *Escherichia coli* str. K-12 substr. W3110 (*E. coli*) was used as an out-of-group sequence. Bootstrap values calculated from 1,000 resampling using neighbor-joining are shown at the respective nodes. Length of the horizontal lines reflects relative evolutionary distances among the sequences. *Sphingomonas* sp. SKA58 should be *Sphingobium* sp. SKA58 on the basis of comprehensive 16S rDNA analysis. γ-HCH degraders are shown in red.

Fig. 3. Patterns for transposition of IS*6100*. Transposition into the same replicon with deletion/resolution (A. intra-replicon 1) or inversion (B. intra-replicon 2), and transposition between two replicons with fusion/integration (C. inter-replicons) are schematically shown. IS*6100* (880 bp) is shown as blue pentagon and squares on the both sides showing direction of the *tnpA* gene (795 bp) and the 14-bp inverted repeats, respectively. All these transpositions cause duplication of IS*6100* and generate 8-bp target (triangle) duplication at the upstream and downstream of IS*6100*. See the reference (Tabata *et al.*, 2016c) for more detailed reaction mechanism.

Fig. 4. Schematic patterns of "editing" roles of IS*6100* found in lindane degraders. Blue and red pentagons indicate IS*6100* and *lin* genes, respectively. Transposition of

IS*6100* on the replicon A (red) (i) into replicon B (blue) with fusion lead to gathering the *lin* genes into one replicon, (ii) into replicon A with deletion lead to trimming unnecessary regions, and (iii) into replicon A with inversion lead to shortening the distance between the *lin* genes.

Fig. 5. Proposed model for the emergence and evolution of lindane-degrading sphingomonad strains. Ancestral various non-lindane-degrading sphingomonad strains with core functions, *e.g.*, LinKLMN-type ABC transporter and β-ketoadipate pathway, turned to be primitive lindane degraders by the acquisition of specific *lin* genes using sphingomonads-specific multiple plasmids and IS*6100*. The primitive lindane degraders were diversified by the involvement of IS*6100* and other mutations. The selective pressure may produce the 'evolved' lindane degraders in the future. See text for detail.