Microsphere (2022) 1(2): 71-81
https://doi.org/10.59118/KRZK6378

Research Article

Correlation of HCH Residues Levels and Metagenomic Lin Protein Sequences at Contaminated Sites

Vivek Negi[1] · Helianthous Verma[2] · Roshan Kumar[3] · Pooja Rani[4] · Amit Kumar Singh[5] · Abhay Bajaj[6] · Rup Lal[7*] ·

Received: 31 July 2022 / Accepted: 27 October 2022 / Published online: 6 November 2022

Abstract

The residue level of hexachlorocyclohexane (HCH) at contaminated dumpsites and nearby or related sites were studied. Water samples and sediments were collected from different HCH contaminated sites for residue analysis. These different sites include solexa data of dumpsite, Ummari, Lucknow (U.P.) (SolexaDS), dumpsite (DS) and two samples at the pond sediments near the pesticide industry, Chinhat, Lucknow (U.P.) (PS1 and PS2). 1km and 5km away from the DS i.e., 1km and 5km respectively were also included in the study. There was a correlation of HCH residue concentration at different sites with LinA and LinB protein distribution. Further, to investigate prevalence of LinA and LinB protein sequences in the metagenomic data sets of sites with different HCH gradients, a gene-centric approach was used. The study showed that in HCH stressed niches, LinA and LinB sequences from the metagenomes were found to cluster, thereby providing evidence in support of the proposed evolution of these proteins in the HCH-contaminated environment.

Key words: HCH, Residue, Lin, Metagenome

Introduction

Since the 1950's, lindane has been used as potential pesticide against crop pests and vector-bone diseases [1]. It is estimated that approximately 4 to 7 million tons of HCH residues have been produced and discharged around the globe in the last six decades following the rise in lindane's application against agricultural pests [2, 3]. The problem rose not only due to γ-HCH (10-12%) which has insecticidal properties but also with the other isomers of HCH which is produced during the production of γ-HCH. Unregulated and unmanaged disposal of other non-insecticidal isomers such as: α- (+/-) (60-70%), β- (10-12%), δ- (6-10%) and ε- HCH (3-4%) had led to formation of many HCH dumpsites all over the world wherever γ-HCH was produced [3, 4]. The α-, β- and γ-HCH isomers were designated as Persistent Organic Pollutants (POPs) in the Stockholm Convention in May 2009 [5] (UNEP 2009). In addition to being the most abundant of the isomers, the α-HCH isomer has also been implicated in carcinogenesis and the β-HCH isomer has been shown to have higher persistence and estrogenic effects [6]. Over the past 20 years of lindane production the industry has utilized inappropriate storage and disposal practices, resulting in widespread environmental HCH contamination and a large number of HCH dumpsites [2]. One of the HCH dumpsites situated in India at Ummari village, Lucknow has been widely studied for characterization of the microbial diversity by culture dependent [7, 8, 9, 10, 11, 12, 13] and culture independent approaches [14, 15] to reveal the potential of microbial community against HCH contamination. These studies reported the presence of efficient HCH degraders and tolerant bacterial species that inhabit under such high HCH stress (450 mg/kg of soil) [14]. Specifically, many of these bacteria shared components of a unique catabolic system for HCH degradation which was named as *lin* system [16, 17, 18, 19, 20].

In the previous study on comparative metagenomic profiling of HCH contaminated soil samples at different gradients i.e. Dump Site (DS, 450 mg/kg of ∑ HCH), one km away from the dumpsite (1KM, 0.7 mg/kg of ∑HCH), five km away from the dumpsite (5km, 0.03mg/kg of ∑HCH) the microbial community dynamics active at such stressed niches were studied. The presence of xenobiotic compounds exerts a selection pressure which has been shown to be linearly correlated to the acquisition of

[1] Ramjas College, Department of Zoology, University of Delhi, Delhi-110007, India. viveknegi2006@gmail.com

[2] Ramjas College, Department of Zoology, University of Delhi, Delhi-110007, India. helianthousverma@gmail.com

[3] Department of Zoology, Magadh University, Bodh Gaya, Bihar – 824234, India. roshanzhc@gmail.com

[4] Department of Zoology, University of Delhi, New Delhi-110007, India. poojaasrivastav@gmail.com.

[5] Department of Zoology, Bhaskaracharya College of Applied Sciences, University of Delhi, Delhi - 110075, India. amitarshifrndz4u@gmail.com.

[6] CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur,440020, Maharashtra, India. abhaybajaj@gmail.com.

Running Title: Correlation of HCH Residues at contaminated sites and Phylogenetic analysis of the Metagenomic Lin protein sequences.

*[7] Author for correspondence: NASI Senior Scientist Platinum Jubilee Fellow. The Energy and Resources Institute, Darbari Seth Block, IHC Complex, Lodhi Road, New Delhi-110003, India, Tel: +91 120 367 9147, Mob 9810785766. Email: ruplal@gmail.com

eISSN: 2583-5327

foreign genes such as the *lin* system genes [14]. Building on our previous study we have investigated and described the prevalence of community metabolic structure in relation to the environmental residue levels of HCH and other contaminants at the dumpsite, 1Km, 5Km away from the DS with pond sediment situated nearby to Indian Pesticide Limited (IPL), Lucknow, former producer of HCH. Moreover, the previous studies were restricted to HCH dumpsites and *lin* gene sequences [18, 19, 21]. In this study, six metagenomic datasets from both HCH dumpsite and IPL industry area made it possible to investigate the effect of HCH pressure on the prevalence and distribution of Lin proteins across the vicinity of the dumpsite as well as pesticide producing industry.

Materials and Methods

a) Sample collection and residue analysis

Soil and water samples were collected in February 2014 (winter session) near the HCH production industrial site [Indian Pesticide Limited (IPL), Industrial Area, Chinhat, Lucknow (U.P.)] Coordinates (26˚54′34.86″N, 81˚04′25.22″E) and HCH dumpsite [Ummari, Lucknow (U.P.)] Coordinates (27˚00′24.7′′N, 81˚08′57.8′′E) along with two more locations 1km Coordinates (27˚00′31.1′′N, 81˚08′54.7′′E) and 5km Coordinates (27˚00′59.5′′N, 81˚08′36.08′′E) distant from dumpsite. Different zones were also selected at both Chinhat industrial area [Pond Sediment (PS), Pond Water (PW) and IPL Drainage Water (IDW)] and Ummari village [DumpSite (DS), one km from DS (1KM), five km from DS (5KM), Water near dumpsite (DSW) and Hand Pump Dumpsite water (HPDS)] **(Figure 1)**. At each zone five replicates of soil samples were collected and then pooled separately to represent that zone. Samples were collected with a sampling auger at depths varying from 0 to 15 cm. Two liters of each of composite water sample were collected from within 200m. All the samples were kept and transported onice to laboratory and the time between sampling and extraction did not exceed 24 h. HCH residues extraction from the samples and subsequent analysis was done using the method described by Prakash *et al.*, 2004 [22]. The physiochemical properties of the sediment samples were analyzed at the Soil Testing Services - Indian Agricultural Research Institute.

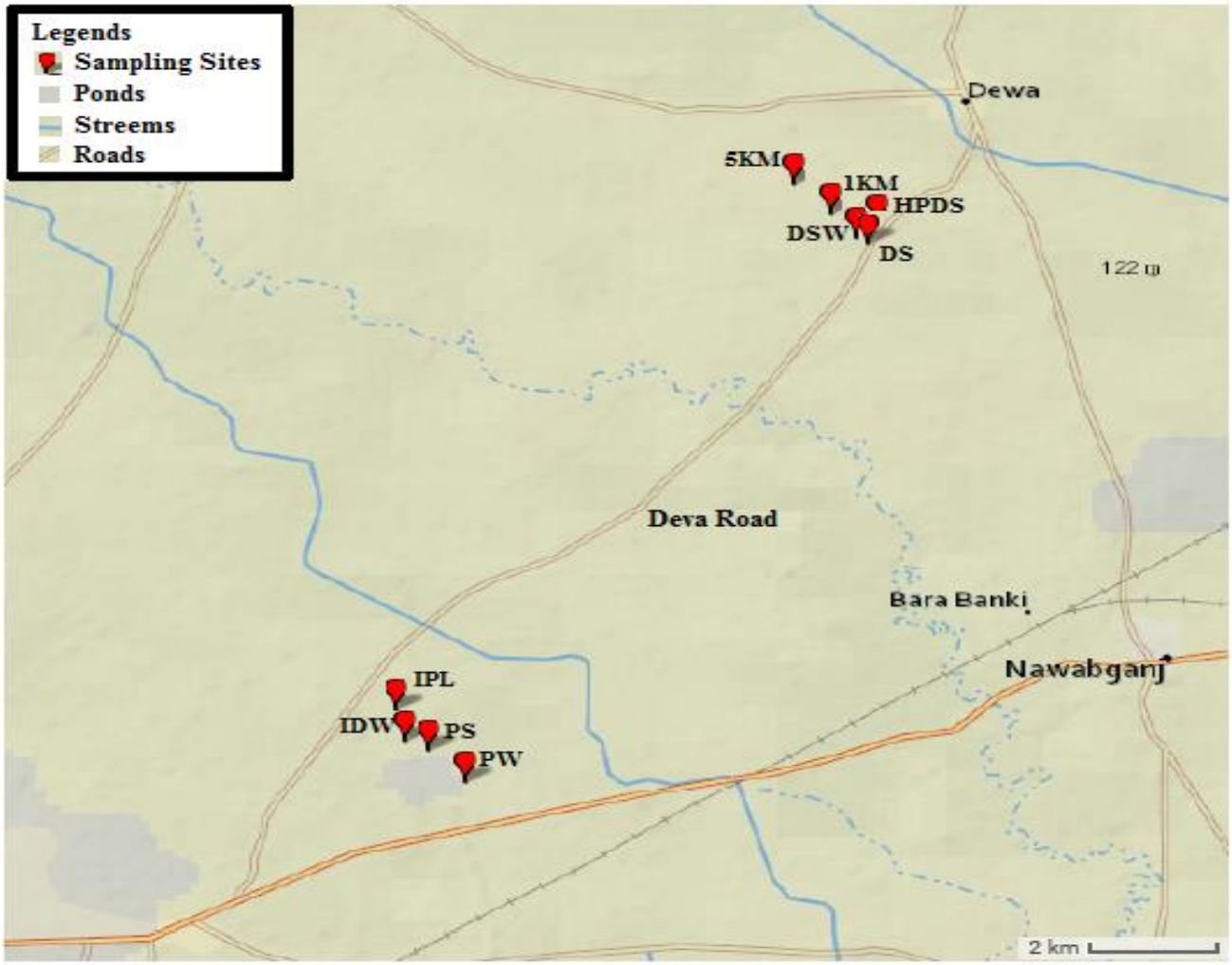

Figure 1: Map of HCH contaminated sampling sites at Ummari Village and Chinhat Village, Lucknow.

b) De-novo assembly of the metagenomes

The metagenomic data was obtained using Illumina HiSeq 2000 platform (Illumina, Inc.) from Pond Sediment1 (PS1) and Pond Sediment2 (PS2). The assembly of sequence was done using Velvet_1.2.03 assembler [23] at k-mer length 39 (expected coverage, auto; insert length, 170 bp; length cutoff, 200 bp and minimum contig standard deviation for insert length, 20 bp) [24]. Optimization of k-mer length was done by running Velvet assembler at different k-mer. Minimum cutoff length for contig was >200bp to ensure better ORFs prediction [25]. Metagenomic data obtained from the 1Km (1KM), 5Km (5KM) distant from the dumpsite and Dumpsite (DS) was assembled using Velvet_1.2.03 assembler at 31 k-mer length, 200bp insert length and "auto" expected coverage [14].

C) Gene centric and phylogenetic analysis

Sequences from PS1 and PS2 (**SRP047092**), DS (**4461840.3**), 1KM (**4461013.3**), 5KM (**4461011.3**) samples were considered for *in silico* analysis of gene and hypothetical proteins or predicted proteins involved in chlorocyclohexane metabolism and phylogenetic analysis of LinA and LinB protein sequences. From metagenomic contigs MetaGene gene finder [26] and MetaProdigal [27] predicted open reading frames (ORFs). KEGG GENES [28] was used to perform pathway annotation of peptide sequences from metagenomic data [(RPS-BLAST) Reverse Position-Specific BLAST; 29 (Bit Score > 60 and bi-directional hit rate (BHR) > 0.95)].

For phylogenetic analysis only the protein sequences having an alignment covering > 80% length of LinA (E-value ≤ 1e-5) and 60% of LinB (E-value ≤ 1e-5) were retained. A maximum likelihood tree with 100 bootstrap replicates was computed with PhyML [30] in which the WAG amino acid substitution model of evolution and four categories of substitution rates were implied.

d) Data availability

The PS data is available at NCBI SRA under the accession no. **SRP047092**. The sequence data of previous metagenomic reads is also available at MG-RAST accessions: 1KM =**4461013.3,** DS =**4461840.3** and 5KM =**4461011.3** (http://www.ebi.ac.uk/ena/data/view/ERP001726).

Results and Discussion

a) Physiochemical and HCH residue analysis across different sites

Among three important nutrients *i.e.*, nitrogen, phosphorus and potassium) nitrogen was the only nutrient that crosses over the permissible limit (120 kg/ha) at all sites. The 5KM soil samples contained high levels of both nitrogen (460 kg/ha) and phosphorus (318 kg/ha). On the other hand, the dumpsite sample contained highest level of potassium (918 kg/ha) as compared to other sites. All the sites have alkaline soil content (pH ≥7) as a result of which the electrical conductivity of the dumpsite sample was found higher (8.5dS/m) due to the saline condition of the soil.

Table 1: HCH Residue (mg/kg) and chemical properties of soil samples.

Samples	PS1	PS2	DS	1KM	5KM
Electrical Conductivity (dS/m)	1.10	0.73	8.50	0.19	0.43
pH	7.72	6.92	7.21	7.81	7.93
Available Phosphorus (P) (kg/ha)	25.3	24.9	60.3	93	318
Organic Carbon(C) (%)	0.85	0.41	30.74	0.45	0.67
Available Nitrogen(N) (kg/ha)	439	502	335	397	460
Available Potassium (K) (kg/ha)	89	90	918	40.5	84.3
Salinity	Normal	Normal	Saline	Normal	Normal
B-HCH (mg/kg)	0.831±0.003	0.833±0.005	182.2±9.7	0.514±0.029	0.295±0.13
α-HCH (mg/kg)	0.074±0.001	0.076±0.002	7.680±4.58	0	0
ΣHCH (mg/kg)	0.905±0.003	0.909±0.006	7.922±8.94	0.514±0.029	0.295±0.013
γHCH (mg/kg)	0	0	35.9±9.7	0	0
δHCH (mg/kg)	0	0	24.1±3.81	0	0

*Values are the mean of three replicates.

Normal salinity levels represented by low concentration of cations (except nitrogen(N)), neutral pH of the samples and electrical conductivity.

Residues of HCH isomers detected in the soil and water samples are mentioned in **Table 1** and **2**. The concentration of all HCH isomers (α-, β-, γ- and δ-HCH) ranged from 0.295±0.013 mg/kg to 7.922±8.94 mg/kg in all soil samples (PS, DS, 1KM and 5KM) and 4.520±0.31 mg/L to 42.16±2.43 mg/L in all water samples (PW, IDW, DSW and HPDS). Specifically, in soil samples, the β-HCH isomer varied from 0.295±0.13 mg/kg to 182.2±9.7 mg/kg whereas in water samples from 4.52±0.31 mg/L to 25.06±2.36 mg/L. In contrast to this, in soil samples, the concentration of α-HCH isomer varied from 0.075±0.001 mg/kg to 7.680±4.58 mg/kg but among water samples, only sample obtained from the dumpsite shown the presence of α-HCH (DSW; 8.08±0.278). Among all the samples, γ- and δ-HCH residual levels were lower than that of α- and β-HCH isomers might be due to their high concentration in t-HCH. HCH residues in soil samples were inversely correlated to the distance from the HCH dumpsite. At the industrial area, the total HCH was greater than 1KM and 5KM sites. The residue level of pond water (PW, 18.520±1.085 mg/L) at IPL was lower than the sample obtained from industrial drainage water (IDW, 42.16±2.43 mg/L).

Table 2: HCH Residue Analysis (mg/L) of water sample collected near to the IPL site and dumpsite.

Samples	PW	IDW	DSW	HPDS
α (g/l)	0	0	8.08±0.278	0
β (g/l)	13±1.000	15.04±1.670	25.06±2.360	4.52±0.310
δ (g/l	5.52±0.095	27.14±0.950	0	0
Σ HCH (g/l)	18.520±1.085	42.16±2.430	33.14±2.559	4.520±0.310

*Values are the mean of three replicates.

Other physiological conditions of the samples, such as, specific conductivity, pH, salinity, and nutrient levels (nitrogen, phosphorus and potassium) are critical to maintain the complex microbial communities at the sites. Among the samples, relatively lower salinity, electrical conductivity, ion concentration of PS2 and PS1 sediment samples indicated a less stressed microbial community with high metabolic efficiency **(Table 1)**. There was significant ($p<0.00001$) differences in electrical conductivity at all the sites, with a maximum at DS (8.5 ds/m). High conductivity might be due to salt ions and high HCH concentration. In addition, with the high conductivity at the dumpsite, high concentration of both potassium and nitrogen was observed. High level of potassium at DS (918 kg/ha) might relate to its bioremediation potential [31] as potassium ions might undergone to exchange reaction with other cations present in the soil, thereby affecting the bioavailability of potassium for plants. Thus saline condition of DS is ideal for potassium retention. In case of nitrogen, high level was found at all the sites might be due to nearby agricultural practices. However, these concentrations may fluctuate with changes soil temperature and moisture. The pH of all the soil samples was alkaline ($pH \geq 7$) and in such condition phosphorus reacts with calcium to form calcium phosphate minerals. Most of the phosphorus in soil is bound by phosphorus minerals in chemical forms, which is then released into the soil as the surface of the soil solution is uncoupled. Thus, phosphorus minerals are low in concentration.

Total HCH from different sites and residue levels of HCH isomers were measured to assess the different levels of HCH contamination. In the previous study, the maximum HCH concentration was observed at the dumpsite and the concentration was found decreasing with increasing distance from the dumpsite [14]. In this study, we have obtained different water and soil samples both from the HCH dumpsite, Ummari village, Lucknow and Indian Pesticide Limited (IPL) a former pesticide factory in Chinhat, Lucknow. The HCH residue analyses of the samples were performed. Among these samples, the residue levels at pond sediment nearby to the IPL industry is not very much evident (> 0.909 mg/kg). Besides organochlorides there are other pesticides [β-endosulfan (220.9ng/mL), α-endosulfan (80.9ng/mL), chlorpyriphos (20.9ng/mL), dimethoate (2.9ng/mL), malathion (1.6ng/mL) and monocrotophos (8.3ng/mL),] present in pond sediment [32]. Also, due to preferential extraction of γ-HCH for its insecticidal property, the residue levels of α- and β-HCH were the highest.

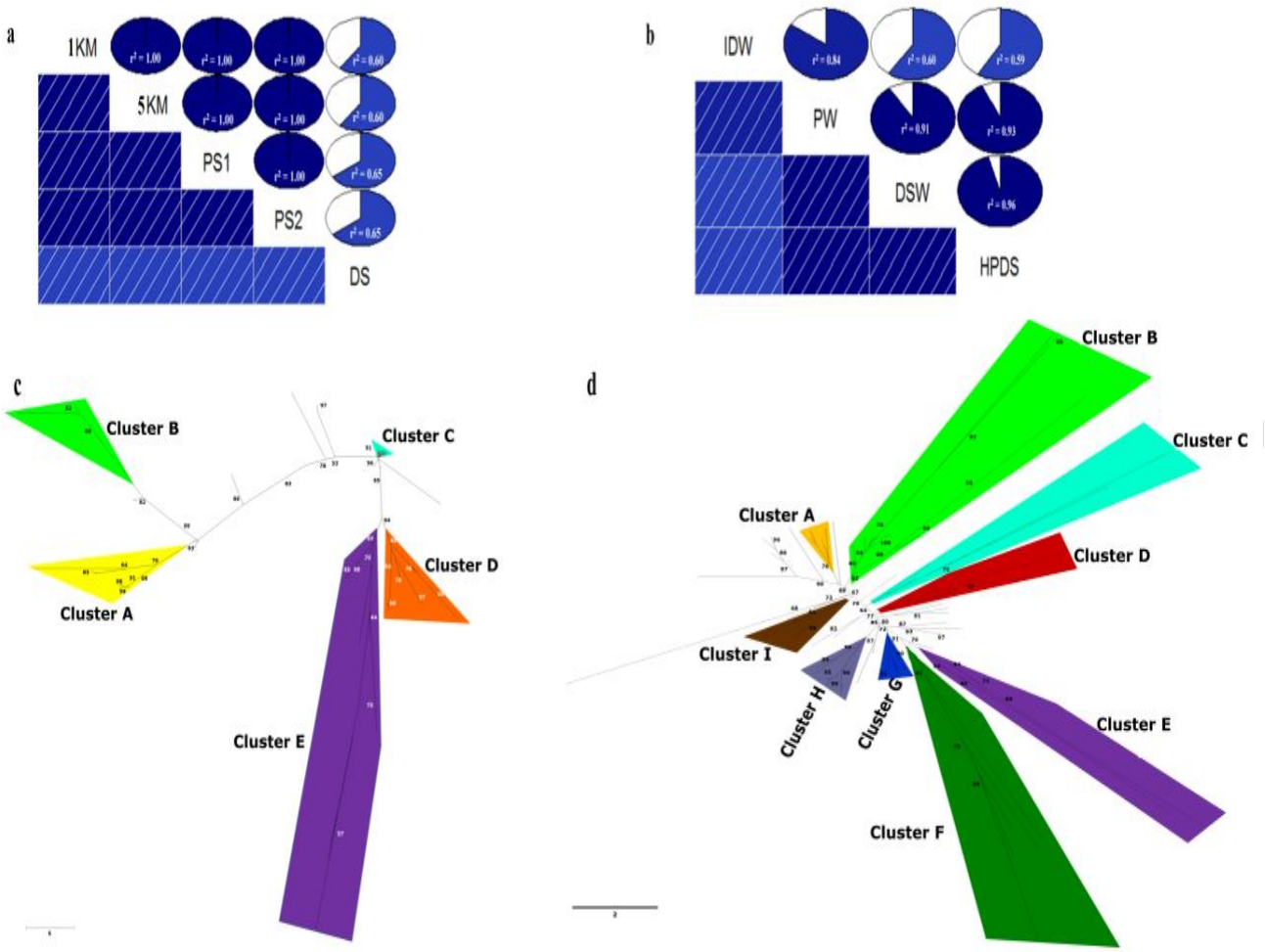

Figure 2: Pearson's r correlation coefficients between HCH isomers for **a.** soil samples (PS1, PS2, DS, 1KM and 5KM) **b.** water samples (PW, IDW, DSW and HPDS). **c.** phylogenetic distribution of LinA proteins in HCH contaminated metagenomes. A distance tree (Maximum-Likelihood) was calculated from the sequences of LinA proteins extracted from PS1, PS2, Solexa DS, DS, 1KM and 5KM. Only protein sequences covering at least 80%, with an E-value ≤ 1e-5 were considered. **d.** phylogenetic distribution of LinB proteins in HCH contaminated metagenomes. A distance tree (Maximum-Likelihood) was calculated from the sequences of LinB proteins extracted from PS1, PS2, Solexa DS, DS, 1Km and 5Km. Only protein sequences covering at least 60% (E-value ≤ 1e-5).

The HCH concentration in samples obtained from IPL industry Chinhat varied from 0.907 mg/kg to 43.0 mg/kg in pond sediment and 18.52 mg/l to 42.16 mg/l in water **(Table 1)**. On the other hand, the residue levels at DS were the highest and showed the correlation with other sample (PS1, PS2, 1KM and 5KM) as, $R^2 \geq 0.60$ **(Figure 2a)**. Interestingly, the residue levels of PS were more correlated ($R^2 = 0.65$) to DS than other samples. Also, in case of water samples, PW was highly correlated ($R^2 > 0.91$) to water samples near DS (DSW and HPDS) **(Figure 2b)**. The β-HCH isomer had the maximum concentration in PS (0.831 mg/kg) and PW samples (13.0 mg/l). Surprisingly, the contamination level of HCH in the water sample which is used for the domestic and agricultural purpose was found greater than the maximum permissible limit (>0.1 mg/l for each isomer and >0.5 mg/l for t-HCH). Hence, this water is unfit for domestic or any other purpose. In addition to this, the prevalence of β-HCH and δ-HCH in PW and PS was observed while α- and γ-HCH are present at relatively lower level or altogether absent. As already stated, α-HCH contains the major portion of HCH muck i.e., 60-70%, but was not identified in pond sediments which suggest that HCH residues in the PW and PS were stored before and now the industry might have discontinued the production of HCH isomers which in turn prevented their further deposition. Hence, only β- and δ- HCH isomers still detected due to their persistence and stable nature.

b) Distribution and diversity of phylogenetic clusters across hexachlorocyclohexane contaminated sites

Using the gene centric approach, several enzymes were unraveled. Enzyme involved in degradation of chlorobenzene and chlorocyclohexane(**ko00361**) was present in all metagenomic datasets (PS1, PS2, Solexa DS, DS, 1KM and 5KM). They are mostly from the family of oxidoreductase, hydrolases, lyases and isomerases. These enzymes help the microbes to tolerate and metabolize different xenobiotic compounds. Among

these, hydroxyquinol 1,2-dioxygenase(**EC 1.13.11.37**) and maleylacetate reductase(**EC 1.3.1.32**) were found enriched in the PS metagenome which catalyzes the conversion of hydroxyquinol to 3-oxoadipate [Hydroxyquinol 1,2-dioxygenase catalyzes the trihydroxylated intermediates (Chlorohydroxyquinol or Hydroxyquinol)] and then maleylacetate reductase converts maleylacetate to 3-oxoadipate. Also, the enzymes after the degradation of chlorinated aromatic compounds [33] yield chlorocatechol. The enzymes might have evolved due to gene duplication and accumulation of multiple mutations. Another enzyme identified predominately was biphenyl-2,3-diol-1,2-dioxygenase (*bph*C) which is an active degrader in number of pathways, including those of phenanthrene , naphthalene [34] and biphenyl [35, 36]. Its has degradation capability for various substrates such as biphenyl -2,3-diol, catechol, 2,3-dihydroxybiphenyl and 3-methylcatechol. Hence, the abundance of these genes determines the functional dynamics of microbes.

Focusing on the *lin* genes, involved in degradation of HCH isomers and are diverging to form several catabolic functions [37], *linA* and *linB* are the two primary genes which acts on the HCH isomers. *linB* belongs to the α/β hydrolase family and encodes for halo alkane dehalogenase whereas *linA* encodes for HCH dehydrochlorinase. The LinA acts on the α- and γ-HCH isomers [16] while *linB* acts on β-HCH, δ-HCH and pentachlorocyclohexanol, yielding a 2,3,4,5,6-pentachlorocyclohexanol (2,3,4,5,6-PCHL), (2,3,4,5,6-PCHL) and tetrachlorocyclohexanediol respectively [19]. The datasets were enriched with sequences of *linA* and *linB* genes (E value > 1e-5) as well as *linC* (*Dehydrochlorinase*), *linR (*LysR family transcriptional regulator*), linD* (*Reductive dechlorinase*), *linF* (*Maleylacetate reductase*), *linGHJ (Acyl-CoA transferase, Acyl-CoA transferase , Thiolase)* and *linKL (Putative ABC transporter system, Putative ABC transporter)* to a lesser extent (Evalue > 1e-5). To investigate the distribution and evolution of LinA, and B proteins in all the metagenomic dataset present at HCH contaminated sites [21], phylogenetic analysis of protein sequences was performed. Phylogenetic analysis led to formation of several clusters of protein sequences based on similarity index. This suggested that LinB proteins were most enriched in DS (20.69%) and LinA proteins in SolexaDS (31.57%). Based on the phylogenetic analysis, cluster C was the most common and most abundant cluster in LinA protein phylogeny, comprising of 26.32% of all LinA homolouges **(Figure 2c)**. Similarly cluster B in LinB protein phylogeny **(Figure 2d)** shows 22.81% of all homologues. Cluster B and F were the most common LinB cluster among the metagenomic datasets, whereas the most abundant cluster was cluster B. The enrichment of LinA protein varies comprehensively among the metagenomic datasets **(Supplementary Table 1)** in which it was found maximum in Solexa DS which could relate to its residual level (7,922±8.94 mg/kg HCH) present at DS **(Table 1)**. On the other hand, the enrichment of LinB protein was evenly distributed among all the metagenomic datasets, independent of the residue levels present **(Supplementary Table 2)**. These findings suggest that the over-representation of Lin proteins in one metagenomic dataset may be due to selection by the environment. To investigate the distribution of LinA and LinB proteins in different HCH contaminated microbiota the protein coding genes derived from metagenomic datasets (PS1, PS2, Solexa DS, DS, 1KM and 5KM) were subjected to phylogenetic analysis. LinA protein sequences are enriched in Solexa DS metagenome (31.57%). ClusterD of LinA is most prevalent in Solexa DS associated metagenome. Simultaneously clusterA and clusterE of DS and 1KM, respectively, are also enriched with LinA protein sequences. More LinA copies might have been recovered from the Dump Site due to the higher HCH residue levels at DS (7,922±8.94 mg/kg). According to the phylogenetic analysis, the protein sequences show diversity and evolution not only at the dumpsite but at nearby sites (1KM) also. Residue levels at PS (<0.909±0.006 mg/kg) are greater than 1KM (0.514±0.029 mg/kg) from the dumpsite. Despite the HCH dumpsite where pressure of HCH is very high, LinA proteins are more clustered at 1KM. This is predicted to be due to the transfer of microbes from DS to 1KM. These microbes may transfer the sequence through horizontal gene transfer (HGT). In case of LinB protein sequences, cluster B is most prevalent at DS and 5KM from dumpsite. High copy numbers of LinA, LinB proteins are present at high residue levels of HCH such as DS (7,922±8.94 mg/kg) but there is no such correlation [21]. The high residue level is not the only factor: other biological factors like HGT by IS elements and plasmids may be responsible for the high copy number. Broad category of substrates of halogenated compounds are preferred by LinB whereas LinA is very restricted to PCCH products, γ-, α-, and δ-HCH [18]. It is due to HCH pressure that LinA proteins are evolving to a greater extent at the DS. LinB, which is evolutionary more advanced and evolved, is evenly distributed at all the contaminated sites [18, 19, 21].

Better correlations can be drawn for Lin protein sequences in different environments with the collection of more metagenomic datasets. Large number of metagenomic datasets can rightfully justify the diversity and distribution of protein sequences and their evolutionary profile in different environment. Along with metagenomic sequencing, the temporal dynamics at such sites will allow for a more comprehensive study regarding evolution of proteins involved in degradation of chlorocyclohexane compounds.

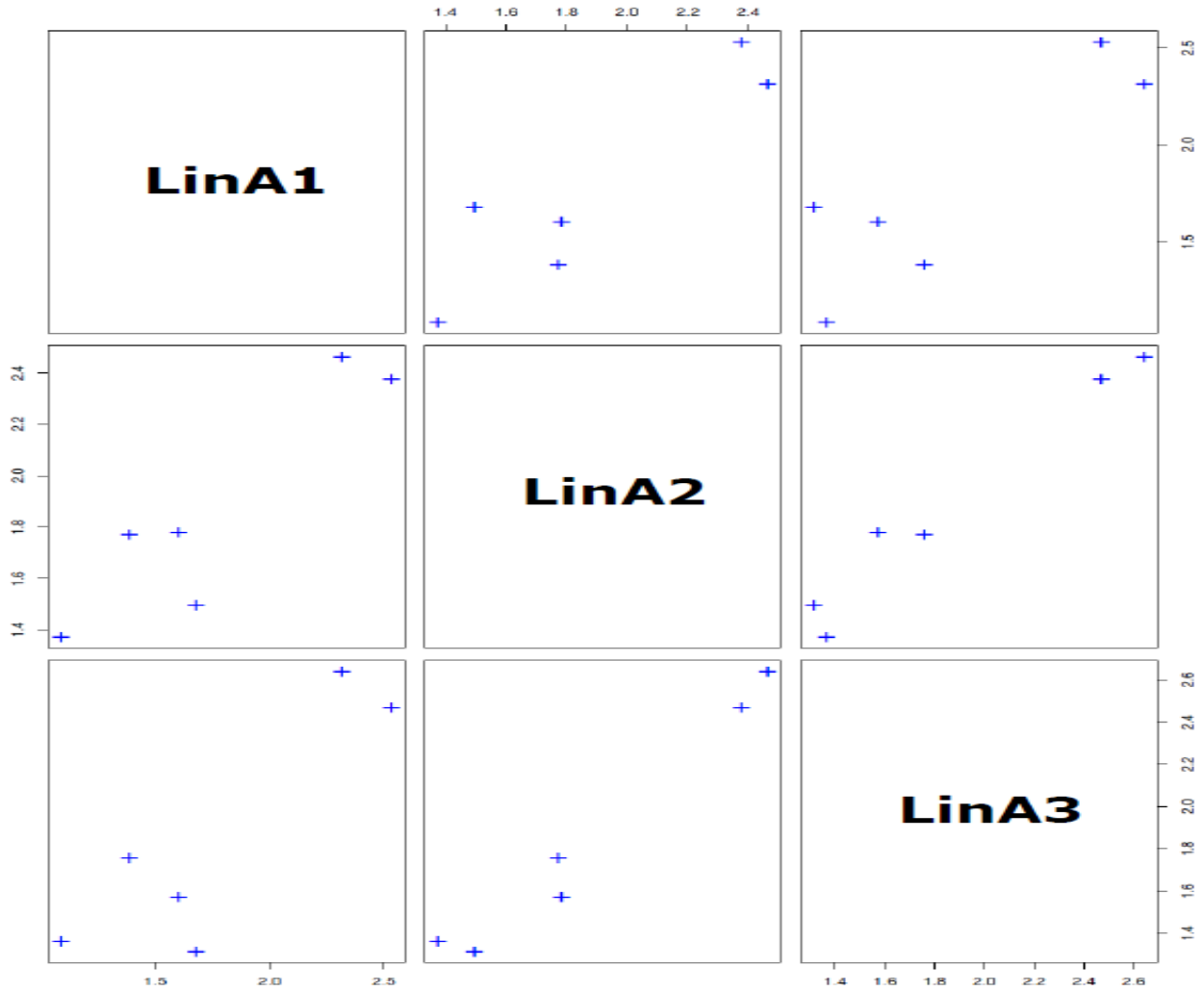

Figure 3: Relative shannon-diversity index based upon the three different genotypes of LinA i.e., LinA1, LinA2 and LinA3. Each dot represents the diversity of the gene with respect to the metagenome: PS1, PS2, Solexa DS, DS, 1KM and 5KM.

C) Diversity and abundance of linA variables across different HCH concentration

To reveal the relative abundance and sequence diversity of *linA* (HCH dehydrochlorinase) and *linB* (haloalkane dehalogenase) gene, their protein sequences were retrieved from the metagenome of the sites under study. This was followed by the Shannon-diversity index analysis to assess their relative abundance with respect to the sites **(Figure 3)**. The highest abundance and diversity of HCH dehydrochlorinases (LinA) and haloalkane dehalogenases (LinB) was found in metagenome of HCH dumpsite samples might corresponds to the maximum concentration of HCH among the sites under study (Table1). Comparatively less amount of these genes are abundant in all the other sites which formed a closed grouping. Focusing on *linA* gene, in which three genotypes are known i.e., *linA1*, *linA2* and *linA3* [38, 13], *linA3* was found most abundant in the DS samples. Notably, *linA3* is a hybrid of *linA1* and *linA2* and was first identified in *Sphingobium* sp. HDIPO4 which was isolated from the same HCH dumpsite [13]. Further, a thermostable *linA* gene copy [39] with these three genotypes and close identity was observed with *linA3* (99.35%) *i.e.*, only a single amino acid substitution of L64I (Leucine with Isoleucine at 64 a.a position) which is however a conservative mutation. This suggests that *linA3* might have the thermostable properties which might be the probable reason of its high abundance at the dumpsite. In addition to this, *linA1* was comparably more abundant in 1KM. From the literature, preferential metabolism of (+)α-HCH by *linA1* and thermostable *linA* copy over (-)α-HCH is already known [38, 39]. Also, higher concentration of (+)α-HCH in environmental samples was reported in the previous studies [40]. This demarcates that both *linA1* and *linA3* contributing in the preferential degradation of (+) α-HCH at the dumpsite which might be present in higher concentration. Except these three datasets (Solexa DS, DS and 1KM), all the other three sites have shown nearly similar abundance of these *linA* genotypes. Hence, linA1 and linA3 genotypes were found to the most abundant at the sites contaminated with HCH. Similarly, in case of *linB*, the gene copy was found most

enriched at the dumpsite followed by the other four sites. Therefore, even the pond sediment is closely situated to IPL industry, but the enrichment of both hydrochlorinases and haloalkane dehalogenases was less may be due to less concentration of environmental pollutant compared to the HCH dumpsite. Hence this suggests that open sinks in which HCH muck were dumped have both the more HCH concentration and microbial community to adapt against the environment rather than the sites present nearby to the HCH producing industry or industrial effluents.

Conclusion

To draw a comparison between the HCH contaminated dumpsite and areas nearby to the organochlorine producing industries, HCH residue at the dumpsite (DS), 1KM, 5KM away from the DS and pond sediment (PS) were assessed. Also, comparative functional profiling of microbial community was analyzed between DS, 1KM, 5KM and PS with respect to the level of HCH contamination. Besides HCH there are many aromatic compounds and heavy metals which get discharged from the industry hence, genes encoding for aromatic compound degradation such as hydrolases, oxidoreductase, isomerases and lyases were identified in the pond sediment metagenome. In addition, enzymes like *hydroxyquinol* 1,2-*dioxygenase*, *maleylacetate reductase* and *biphenyl*-2,3-*diol*-1,2-*dioxygenase* (*bphC*) specifically degrading chlorocyclohexane and chlorobenzene have been discussed in detail. The diversity and distribution of LinA protein sequences were found abundant in response to HCH contamination stress. Whereas in case of LinB protein sequences the distribution was even throughout the sites. Hence, this information was retrieved by high-throughput sequencing of complex communities, with residue and phylogenetic analysis which provided a clear picture of microbial community and gene enrichment with response to environmental contaminants and their concentration. To further extend this work, it is critical to acquire knowledge of complex microbial metabolism in the context of the environmental conditions in which they are evolving.

Competing Interests

The authors have declared that they have no competing or conflicting interest.

Acknowledgments

The work was supported by grants from the Department of Biotechnology (DBT), Government of India, under project (BT/PR3301/BCE/08/875/11), the All India Network Project on Soil Biodiversity-Biofertilizers (ICAR) XIth Plan (12-A/2007-lA ll) and DU-DST Promotion of University Research and Scientific Excellence (PURSE) grant. V.N., R.K. and H.V. gratefully acknowledge the Council for Scientific and Industrial Research (CSIR) and the National Bureau of Agriculturally Important Microorganisms (NBAIM) NBAIM/AMASS 2014–17/PF/9/BG/378 for providing research fellowships.

References

1. Busvine JR, Pal R. 1969. The impact of insecticide-resistance on control of vectors and vector-borne diseases. Bull World Health Organ. 40:731-44.

2. Jit S, Dadhwal M, Kumari H, Jindal S, Kaur J, Lata P, Niharika N, Lal D, Garg N, Gupta SK, Sharma P, Bala K, Singh A, Vijgen J, Weber R, Lal R. 2010. Evaluation of hexachlorocyclohexane contamination from the last lindane production plant operating in India. *Environmental science and pollution research international.* 18:586-597. doi: 10.1007/s11356-010-0401-4.

3. Vijgen J, Abhilash PC, Li YF, Lal R, Forter M, Torres J, Singh N, Yunus M, Tian C, Schäffer A, Weber R. 2010. HCH isomers as new Stockholm Convention POPs are we on the way to manage lindane (γ-HCH), and waste HCH isomer (α- and β-HCH) pollution around the world? *Environmental science and pollution research.* 5:363–393.

4. Willet KL, Utrich EM, Hites RA. 1998. Differential toxicity and environmental facts of hexachlorocyclohexane isomers. *Environmental science & technology*. 32:2197–2207. doi: 10.1021/es9708530.

5. UNEP. 2009. Report of the Conference of the Parties of the Stockholm Convention on Persistent Organic Pollutants on the work of its fourth meeting. UNEP/POPS/COP.4/38, 2009.

6. EPA. 2003. Five-year review fact sheet, Geigy Chemical Corp. Site, Aberdeen, Moore County, North Carolina, Oct. 2003. US EPA.

7. Singh A, Lal R. 2009. *Sphingobium ummariense* sp. nov., a hexachlorocyclohexane (HCH)-degrading bacterium, isolated from HCH contaminated soil. *International Journal of Systematic and Evolutionary Microbiology*. 59:162–166. doi:10.1099/ijs.0.65712-0.

8. Dadhwal M, Jit S, Kumari H, Lal R. 2009. *Sphingobium chinhatense* sp. nov., a hexachlorocyclohexane (HCH) degrading bacterium isolated from an HCH dump site. *International Journal of Systematic and Evolutionary Microbiology*. 59:3140-3144. doi: 10.1099/ijs.0.005553-0.

9. Bala K, Sharma P, Lal R. 2010. *Sphingobium quisquiliarum* sp. nov., P25T a hexachlorocyclohexane (HCH) degrading bacterium isolated from HCH contaminated soil. *International Journal of Systematic and Evolutionary*

Microbiology. 60:429-433. doi: 10.1099/ijs.0.010868-0.

10. Sharma P, Verma M, Bala K, Nigam A, Lal R. 2010. *Sphingopyxis ummariensis* sp. nov., isolated from hexachlorocyclohexane (HCH)-dumpsite in north India. *International Journal of Systematic and Evolutionary Microbiology*. 60:780-784. doi: 10.1099/ijs.0.008805-0.

11. Kumar R, Dwivedi V, Negi V, Khurana JP, Lal, R. 2013. Draft Genome Sequence of *Sphingobium lactosutens* Strain DS20 isolated from a Hexachlorocyclohexane (HCH) Dumpsite. *Genome Announcement.* 1:e00753-13. doi: 10.1128/genomeA.00753-13.

12. Negi V, Lata P, Sangwan N, Das S, Gupta SK, Rao DLN, Lal R. 2014. Draft genome sequence of hexachlorohexane (HCH) degrading *Sphingobium lucknowense* strain F2T isolated from the HCH dumpsite. *Genome Announcement.* 2:e00788-14. doi:10.1128/genomeA.00788-14.

13. Verma H, Kumar R, Oldach P, Sangwan N, Khurana JP, Gilbert JA, Lal R. 2014. Comparative genomic analysis of nine *Sphingobium* strains: insights into their evolution and hexachlorocyclohexane (HCH) degradation pathways. *BMC Genomics*. 15:1014. doi: 10.1186/1471-2164-15-1014.

14. Sangwan N, Lata P, Dwivedi V, Singh A, Niharika N, Kaur J, Anand S, Malhotra J, Jindal S, Nigam A, Lal D, Dua A, Saxena A, Garg N, Verma M, Kaur J, Mukherjee U, Gilbert JA, Dowd SE, Raman R, Khurana P, Khurana JP, Lal R. 2012. Comparative metagenomic analysis of soil microbial communities across three hexachlorocyclohexane contamination levels. *PLoS One*. 7: e46219. doi: 10.1371/journal.pone.0046219.

15. Sangwan N, Verma H, Kumar R, Negi V, Lax S, Khurana P, Khurana JP, Gilbert JA, Lal R. 2013. Reconstructing an ancestral genotype of two hexachlorocyclohexane-degrading *Sphingobium* species using metagenomic sequence data. *The ISME journal.* 8:398–408. doi: 10.1038/ismej.2013.153.

16. Nagata Y, Ohtomo R, Miyauchi K, Fukuda M, Yano K, Takagi M. 1994. Cloning and sequencing of a 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase gene involved in the degradation of gamma-hexachlorocyclohexane in *Pseudomonas paucimobilis*. *Journal of Bacteriology*. 176:3117-3125.

17. Nagata Y, Endo R, Ito M, Ohtsubo Y, Tsuda M. 2007. Aerobic degradation of lindane (γ-hexachlorocyclohexane) in bacteria and its biochemical and molecular basis. *Applied and Environmental Microbiology*. 76:741–752. doi: 10.1007/s00253-007-1066-x.

18. Kumari R, Subudhi S, Suar M, Dhingra G, Raina V, Dogra C, Lal S, van der. Meer JR, Holliger C, Lal R. 2002. Cloning and characterization of *lin* genes responsible for the degradation of hexachlorocyclohexane isomers by *Sphingomonas paucimobilis* strain B90. *Applied and Environmental Microbiology*. 68:6021–6028. doi: 10.1128/AEM.68.12.6021-6028.2002.

19. Sharma P, Raina V, Kumari R, Malhotra S, Dogra C, Kumari H, Kohler HPE, Buser HR, Holliger C, Lal R. 2006. Haloalkane dehalogenase LinB is responsible for beta- and delta-hexachlorocyclohexane transformation in *Sphingobium indicum* B90A. *Applied and Environmental Microbiology*. 72:5720–5727. doi: 10.1128/AEM.00192-06.

20. Lal R, Pandey G, Sharma P, Kumari K, Malhotra S, Pandey R, Raina V, Kohler HPE, Holliger C, Jackson C, Oakeshott JG. 2010. Biochemistry of Microbial Degradation of Hexachlorocyclohexane and Prospects for Bioremediation. *Microbiology and Molecular Biology Reviews*. 74:58-80. doi: 10.1128/MMBR.00029-09.

21. Lal D, Jindal S, Kumari H, Jit S, Nigam A, Sharma P, Kumari K, Lal R. 2013. Bacterial diversity and real-time PCR based assessment of *linA* and *linB* gene distribution at hexachlorocyclohexane contaminated sites. *Journal of Basic Microbiology*. 53:1-11. doi: 10.1002/jobm.201300211.

22. Prakash O, Suar M, Raina V, Rawat C, Pal R, Lal R. 2004. Residues of hexachlorocyclohexane (HCH) isomers in soil and water samples from Delhi and adjoining areas. *Current Science*. 87:73-77.

23. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 18:821–829. doi: 10.1101/gr.074492.107.

24. Negi V, Lal R. Metagenomic Analysis of a Complex Community Present in Pond Sediment. J Genomics. 2017 Mar 10;5:36-47.

25. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes, *Nature Biotechnology*. 31:533-538. doi: 10.1038/nbt.2579.

26. Noguchi H, Park J, Takagi T. 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*. 34:5623-5630. doi: 10.1093/nar/gkl723.

27. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics*. 28:2223-2230. doi: 10.1093/bioinformatics/bts429.

28. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research*. 32:D277–D280. doi: 10.1093/nar/gkh063.

29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215:403–410.

30. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*. 59:307-321. doi: 10.1093/sysbio/syq010.

31. Walker A, Jurado-Exposito M, Bending GD, Smith VJR. 2001. Spatial variability in the degradation rate of isoproturon in soil. *Environ Pollut*. 111:407–415. doi: 10.1016/S0269-7491(00)00092-0.

32. Anjum R, Malik A. Evaluation of mutagenicity of wastewater in the vicinity of pesticide industry. Environ Toxicol Pharmacol. 2013; 35: 284–291.

33. Chatterjee DK, Kellogg ST, Watkins DR, Chakrabarty AM. 1981. Molecular Biology, Pathogenicity, and Ecology of Bacterial Plasmids. In: Levy SB, Clowes RC, Koenig EL, editors. *Plasmids in the Biodegradation of Chlorinated Aromatic Compounds*. Springer US. p519-528.

34. Denome SA, Stanley DC, Olson ES, Young KD. Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway. J Bacteriol. 1993 Nov;175(21):6890-901.

35. Kimbara K, Hashimoto T, Fukuda M, Koana T, Takagi M, Oishi M, Yano K. Cloning and sequencing of two tandem genes involved in degradation of 2,3-dihydroxybiphenyl to benzoic acid in the polychlorinated biphenyl-degrading soil bacterium Pseudomonas sp. strain KKS102. J Bacteriol. 1989 May;171(5):2740-7.

36. Hofer, B., Eltis, L. D., Dowling, D. N. & Timmis, K. N. (1993). Genetic analysis of a Pseudomonas locus encoding a pathway for biphenyl/ polychlorinated biphenyl degradation. Gene 130, 47–55.

37. Lal R, Dogra C, Malhotra S, Sharma P, Pal R. 2006. Diversity, distribution and divergence of *lin* genes in hexachlorocyclohexane-degrading *sphingomonads. Trends in Biotechnology*. 24:121-130. doi: 10.1016/j.tibtech.2006.01.005.

38. Suar M, Hauser A, Poiger T, Buser HR, Mu¨ller MD, Dogra C, Raina V, Holliger C, van der Meer JR, Lal R, Kohler HP. 2005. Enantioselective Transformation of alpha-Hexachlorocyclohexane by the Dehydrochlorinases LinA1 and LinA2 from the Soil Bacterium *Sphingomonas paucimobilis* B90A. *Applied and Environmental Microbiology*. 7:8514–8518. doi: 10.1128/AEM.71.12.8514-8518.2005.

39. Macwan AS, Kukshal V, Srivastava N, Javed S, Kumar A, Ramachandran R. Crystal Structure of the Hexachlorocyclohexane Dehydrochlorinase (LinA-Type2): Mutational Analysis, Thermostability and Enantioselectivity. PLoS ONE 7(11): e50373. doi:10.1371/journal.pone.0050373.

40. Müller MD, M Schlabach, M Oehme. 1992. Fast and precise determination of α-hexachlorocyclohexane enantiomers in environmental samples using chiral high-resolution gas-chromatography. *Environ. Sci. Technol.* 26:566-569.

Supplementary Material

Supplementary Table 1: Relative abundance of LinA protein among different metagenomes

Samples	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
Pond Sediment (PS1)	-	-	2	2	1
Pond Sediment (PS2)	-	-	4	1	-
Solexa DumpSite (Solexa DS)	4	2	2	6	4
DumpSite (DS)	5	2	3	3	3
One KM from DS (1KM)	1	-	1	-	5
Five KM from DS (5KM)	1	1	3	-	1

Supplementary Table 2. Relative abundance of LinB protein among different metagenomes

Samples	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	Cluster F	Cluster G	Cluster H	Cluster I
Pond Sediment (PS1)	-	2	-	-	2	1	-	3	2
Pond Sediment (PS2)	2	2	1	-	-	1	-	3	-
Solexa DumpSite (Solexa DS)	1	-	-	2	-	2	-	-	-
DumpSite (DS)	-	4	2	2	1	-	2	1	-
One KM from DS (1KM)	1	1	-	1	2	2	1	-	2
Five KM from DS (5KM)	-	4	1	1	1	1	1	1	1