

Microbial Genes Involved in the Aromatic Compound Degradation Deciphered Through Metagenomics Analysis of Industrial Wastewater.

Priti Pandit

Gujarat Biotechnology Research Centre

Raghawendra Kumar

Gujarat Biotechnology Research Centre

Dinesh Kumar

Gujarat Biotechnology Research Centre

Zarna Patel

Gujarat Biotechnology Research Centre

Labdhi Pandya

Gujarat Biotechnology Research Centre

Chaitanya Joshi (✉ director.gbrc@gmail.com)

Gujarat Biotechnology Research Centre <https://orcid.org/0000-0003-4680-2861>

Research

Keywords: Aromatic compounds, Microbial diversity, Common Effluent Treatment Plant, and Shotgun Metagenomics

Posted Date: August 25th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-62394/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Understanding microbial and functional diversity from different stages of common effluent treatment plant (CETP) plays an important role to enhance the treatment performs of wastewater systems. However, unraveling microbial interactions as well as utilization of substrate involved in complex microbial communities is a challenging task. Hence, we demonstrate an integrated approach of shotgun metagenomics and whole genome sequencing to identify the microbial diversity and genes involved in degradation of benzoate, 1,2-dichloroethane and phenylalanine metabolism and degradation pathways from CETP microbiome.

Results: The taxonomy profile was annotated using the Ribosomal Database Project (RDP) database in the MG-RAST server. The results showed that, bacteria accounted for 98.46% was the most abundant domain, followed by Eukaryota (0.10%) and Archea 0.02%. At Phylum level, *Proteobacteria* (28.8%) were dominant, followed by *Bacteroidetes* (16.1%), *Firmicutes* (11.7%) and *Fusobacteria* (6.9%). The most dominated species were *Klebsiella pneumoniae*, *Wolinella succinogenes*, *Pseudomonas stutzeri*, *Desulfovibris vulgaris*, *Clostridium sticklandii*, and *Escherichia coli*. The Clusters of Orthologous Groups (COGs) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, revealed the presence of the genes responsible for the metabolism and degradation of aromatic compounds. This information was validated with the whole genome analysis of the bacteria isolated from the CETP.

Conclusion: The two type of integrated meta omics analyses revealed that the metabolic and degradation capability at both community wide and individual bacterial levels. In addition, we demonstrated that microbial diversity changes with the treatment process in which inlet of CETP effluent shows higher dominancy of *Proteobacteria* whereas in textile industry outlet the high abundance of *Firmicutes* was observed. We foresee this approach would contribute in designing the bioremediation strategies for the industrial treatment process.

Background

The different organic pollutants released by the increasing number of industries is the major concern towards the environmental and health-related problems that have harmful effects on living beings [1]. Henceforth, as a measure for environment safety, the concept of common effluent treatment plants (CETP) is adapted looking to the needs for the treatment of effluent from small-scale industries units collectively [2]. Wastewater generated at small-scale industries is transported to a CETP where it is treated before being discharged into the open environment [3]. Though the CETP have been functioning, the treated effluents discharged from these plants are still coloured [2].

The activated sludge and wastewater from CETP contains highly diverse and dynamic microbial community, which are mainly responsible for the biodegradation of the pollutants in the wastewater systems [4]. However, despite various applications, few studies have been reported on identifying the microbial community for the degradation process in industrial wastewater. Moreover, early studies only focused on the diversity of microbes in a single wastewater system. Hence, there is a need to investigate the microbial diversity and functional structure in complex wastewater systems. Recently, the metagenomics approach has been increasing rapidly which targets the diversity of the microbial community within a defined environment. Metagenomes provide identification of microbial communities based on DNA sequence as well as to predict genes that confer novel functions [5]. Moreover, it also helps to construct novel insights into the metabolic pathways and in the reconstruction of microbial genomes. Metagenomics studies of complex niches can provide a means to assess the structural and functional aspects of the microbial community and bring to light the biochemical capacity associated with an ecosystem [4]. The metagenomics study will enable to identify the functional genes and metabolic pathways in a requisite ecosystem. The metagenomics study will enable to identify the functional genes and metabolic pathways in a requisite ecosystem. Recent, reports available describes the microbial diversity and metabolic pathways via metagenomics for mining of genes for polyhydroxyalkanoates (PHAs) production [6], aromatic degradation pathways [7], new drugs and antibiotics [8] and arsenic degradation [9].

The wastewater generated from the small scale industries contains the complex mixture of the manufacturing of textile dyes, paints, chemicals, solvents, and textile released wastewater [10]. Hence, the microbial community present in such ecosystem are exposed to diverse array of substrates and are responsible for the degradation capacity of these molecules in the target niche. Since each wastewater treatment plant has unique ecological niche as they exhibit a unique bacterial community along with metabolic composition [11]. Therefore, understanding of the microbial community structure of wastewater treatment plant is important to improve the effectiveness and efficiency of treatment of wastewater. Therefore, the goal of the study involved finding out the microbial community present in the textile wastewater via the metagenomics approach. Simultaneously, compare the dominant microbial populations with different textile treatment plants. Additionally, to construct the metabolic pathway and its enzymes for dye degradation.

In this study, the wastewater and active sludge from the different stages of the Vatva CETP were processed for metagenomics analysis as well as conventional culturing. Two potential bacterial isolates were isolated from the aeration tank and primary sludge was also processed for whole-genome sequencing using Ion Torrent Platform. The metagenomics analysis was carried out to understand the taxonomical and functional potential of the microbiome in the targeted niche. Further, an in-depth comparison of CETP metagenomics with whole-genome sequencing was carried out for the functional validation. Additionally, results obtained from the characterization of the effluent were explored

for the benzoate degradation, 1,2-dichloroethane degradation, and phenylalanine metabolism and degradation pathway encoding gene sequences.

Methods

Sampling

The samples (water and sludge) were collected from different stages of Common Effluent Treatment Plant (CETP), Vatva, Ahmedabad, Gujarat, India that receives effluent from approximately 674 different industrial units like the textile industry, chemicals, dye, and dye intermediate manufacturing units, etc. The CETP treats approximately 1,60,000 m³ of effluent per day. Details of the sampling points are given in the (Fig. 1). All the samples were collected in 2-liter capacity sterile containers in triplicates. Samples were then transported to the laboratory in a portable refrigerator and stored at 40 °C until further processing. The samples are designated as follows: TP01: CETP collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air Flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal.

Physicochemical and GC-MS analysis

Physicochemical analysis of water and sludge samples were carried out to explore the external environmental factors responsible for microbial diversity. To determine the total organic carbon (TOC), the water and sludge samples were filtered through a 0.22 µm membrane filter and then analyzed with a TOC analyzer (Shimadzu TOC-VCSN/TNM-1, Japan). For heavy metal (Cr, Zn, Cu, Ni, and Fe) were extracted from the filter samples as per the standard procedure and measured using inductively coupled plasma mass spectrometer (ICP-MS) (Perkin 3300, USA). The Heavy metal analysis was done at Sophisticated Instrumentation Centre for Applied Research Testing (SICART) Anand, Gujarat, India. For GC/MS analysis, 50 mL of sample was taken out from each sampling stage and extracted with 50% of dichloromethane. The extract was combined, dehydrated with anhydrous sodium sulphate, and concentrated using a rotary evaporator. The residue was dissolved in 1.0 mL of dichloromethane (chromatographically pure grade) and filtrated with a 0.22µm filter. GC/MS analysis was performed on Perkin Elmer with Elite-5MS capillary column (30 mm x 0.25mm) was used. Helium serves as carrier gas at a flow rate of ml min⁻¹. The MS spectra were set from 100 - 650 m/z and analysed in ESI mode.

Metagenomics DNA extraction and library preparation

Total metagenomics DNA was extracted from water-using power water DNA isolation kit power and sludge sample using soil DNA isolation kit (MoBio Laboratories Inc., CA, USA), respectively by following the manufacturer's protocol. Prior, to extraction the water sample was filtered and the sludge sample was centrifuged 8000×g 10 min, and the pellet was used for DNA isolation. Quality and quantity of total DNA was assessed using agarose gel electrophoresis and Qubit dsDNA High Sensitivity assay kit with Qubit 4 fluorometer and stored at -20°C until further processing. The metagenome shotgun libraries were prepared using the Ion Xpress™ Plus fragment library kit (ThermoFisher Scientific, USA). The purified libraries were quantified before pooling using Qubit 4 fluorometer (Invitrogen, USA) and sequencing was carried out using 530 chip, with 400 bp chemistry on the Ion S5 platform.

Data processing and analysis

MG-RAST server version 4.0.3 was used for the reads pre-processing and taxonomic annotation of the sequences. Low-quality sequences were filtered using the default parameters in the MG-RAST serve. The obtained raw sequencing dataset and post-processing quality-filtered sequence information is provided in Table 2 respectively. For the assessment of taxonomic profiles, the reads having a BLAST hit with an e-value of 1e-5, a minimum identity of 60%, the minimum abundance of 10 and a minimum alignment length of 50bp using the Ribosomal Database Project (RDP) (Meyer et al., 2008). The functional annotations of the reads were carried out using SqueezeMeta analysis pipeline at <https://github.com/jtamames/SqueezeMeta>.

Whole genome Sequencing and analysis

Isolation, DNA extraction and whole genome sequencing

The bacteria were isolated from the same samples collected from CETP, Vatva. Dye decolorization bacteria were isolated in half strength nutrient agar with 10 mg/100ml dyes collected from Jetpur, Gujarat. The plates were then incubated at 37° C. The samples which got decolorization zone were further isolated/subcultured to obtain pure cultures. A total of four different bacteria were isolated from plates which showed decolorization zone on the plate. The genomic DNA of these dye decolorization bacteria was isolated using bacterial purlink genomics DNA isolation Kit as per manufacturer instructions. Cultures were identified by sequencing of its 16S rRNA gene. A BLASTn search against the NCBI database was carried out to find the similar sequences in the database. The genomic DNA was used for preparing shotgun libraries and whole genome were sequenced as mentioned as earlier.

Post-sequencing and analysis

The genome sequencing of four potential bacterial isolates was carried out using Ion Torrent S5 Platform. Low quality sequence data were filtered and trimmed using the FastQC V0.10.1. The clean reads were then assembled using SPAdes version 3.9.0 with different k-mer length of k21, k33, k55, k77, k99, k127 and QCAST (Quality Assessment Tool for Genome Assemblies) was used to evaluate the quality of assembly. The annotation was carried out using Annotate Assembly and Re-annotate Genomes with Prokka annotation pipeline v1.12 version and Rapid Annotation Subsystem Technology (RAST) version 2.0 (<http://rast.theseed.org/>). The total length of *Pseudomonas aeruginosa* genome strain was 6,356,997 bp, with a GC content of 65.5 % and an N50 value of 9167 bp with 6,021 coding sequences, 62 tRNA, 2 rRNA and 1 CRISPR genes, while in *Bacillus licheniformis* the total length was 4,364,932 bp with a GC content of 47.0% and an N50 value of 711073 with 4373 coding sequence, 67 tRNA and 7 rRNA. The degradation and metabolic pathways as well as key genes responsible to degrade aromatic compounds and its intermediates were examined using the NCBI and KEGG databases.

Statistical analysis

The parameters were analysed in triplicate and represented as mean \pm standard deviation. The one-way analysis of variance (ANOVA) was applied to assess treatment differences between physicochemical parameters using the SPSS version 25.0 significance between means was tested by DMRT's Test ($p \leq 0.05$). Species richness estimators and community diversity indices were calculated using PAST software version 3.20 (Hammer and Harper 2001). The abundance of the KEGG and COGs functions were used for clustering of the samples following the Principal component analysis (PCA). The PCA analysis and heat map construction was performed by using Microbiome Analyst [12].

Results

Physicochemical and GC-MS analysis

The spatial heterogeneity in physicochemical parameters was revealed in the different sampling points (Table 1). The heavy metal concentration such as iron varied significantly ($p=0.05$) between the sampling points. Iron concentration was higher (1.42 mgkg^{-1}) in the collection tank (TP01) as compared to the discharge of final effluent (TP06) (0.324 mgkg^{-1}). The other heavy (Cr, Ni and Zn) decreased from 1.09 mgkg^{-1} to 0.195 , 0.10 to 0.07 mgkg^{-1} and 0.056 to 0.041 mgkg^{-1} respectively. Analysis of variance results show that there was significant difference in chromium concentration and the mean value of chromium was significantly different in all sampling point except TP06 and TP05 ($p < 0.05$). As effective as chemical, physical and biological treatment reduces the load. Whereas, the concentration of copper was slightly higher in the final effluent; TP06 (1.53 mgkg^{-1}) than the collection tank; TP01 (0.23 mgkg^{-1}). However, the permissible limit of the heavy metals was in the range as prescribed by the central pollution control board for the treated effluent standards (CPCB,2012). Variation in the TOC level was observed in different stages of the treatment process. The TOC level was significantly higher in TP07 (414.35 ppm) and TP03 (178.8 ppm) and reduced dramatically in TP08 and TP06 after treatment. The total organic carbon showed the difference between the different sampling points and suggested their contribution to the diversity of the microbial community in the CETP.

The GC-MS analysis was carried out for the characterization of the compounds present in the CETP, vatva (Table S1). On comparison of the mass spectra of the constituents with the NIST library, different chemicals compounds were identified in effluent treatment plant. The compounds identified in the CETP consisted of a complex mixture of phthalate, phenolic compounds, aromatic and aliphatic carboxylic acids, amines and alkanes. In the TP01 most abundant compounds were identified as benzensulphonamide (28.56%), 1-hexyl-2-nitrocyclohexane (11.17%), oxalic acid (9.71%), cyclopentane (5.85%), phthalic acid (5.47%) and bis(2-ethylhexyl) (2.81%) followed by 2-methylhexacosane (2.01%), 3-ethyl-3-methyl (2.05%) and 1,2 benzenedicarboxylic acid (2.00%). Compounds like bis(2-ethylhexyl) phthalate, cyclopentane, (2-hexyloctyl), 2-methylhexacosane, decane, 3-ethyl-3-methyl, octatriacontane, 1,3-dibromo, benzenesulfonamide and 4-amino-n-ethyl were not detected in the TP06 sample. Moreover, in TP01 and TP06 samples showed the elevated peak area of 1,2 benzenedicarboxylic acid (3.72%) and 1-hexyl-2-nitrocyclohexane (8.34%) levels over the other TP02, TP03, TP04 & TP05 samples. Among all sampling points, TP02 and TP07 shows the presence of the amines such as dimethoxyamine, 1-butanamine, 1-propanamine and 1,1'-biphenyl]-4,4'-diamine, 2,2'-dichloro. The maximum percentage of the oxalic acid with peak area 28.59% was shown by the TP08 followed by the TP04 (21.09%). Moreover, in sludge samples (TP07&TP08) different compounds such as benzeneacetic acid, 2-undecene, dichloroacetic acid and cyclohexane were detected (Table S1).

Microbial community structure and taxonomic profile

The characteristic features and annotation of metagenome datasets were summarized in Table 2. The rarefaction curve was obtained to determine whether the level of sequencing carried out in a sample is sufficient to represent its true diversity. The rarefaction curve was plotted between species richness and the number of expected OTUs obtained per sample (Fig. S1). Rarefaction curves almost reached an asymptote state indicating that full community coverage was achieved which were reflected by a higher Chao1 index which ranged from 570.80 to 1018 (Table 4). The alpha diversity was estimated to determine the species richness and diversity of the given system. The different diversity indices

were estimated which showed that the microbial diversity varies slightly with the treatment stages and has not reduced the microbial abundance within the stages as revealed by Simpson diversity indices (Table 4). Furthermore, Simpson's index of diversity (1-D) ranges from 0.77 to 0.90 and Shannon's diversity index ranges from 3.05 to 3.95 was relatively higher, which reflects the richness in diversity. At CETP, the community sampled from TP06 was less diverse as compared to other samplings points. Whereas, in TP07 the highest level of richness and diversity was observed. Other diversity indices and data from individual samples were elevated. TP02 showed relatively higher Shannon's Diversity Index (3.95) which disperse the increase in the species richness and evenness and hence the diversity in TP02 also increases as compared to other sampling points.

The taxonomic diversity of the microbial community was carried out by using the RDP classifier based on 16S rRNA gene sequences in the shotgun metagenome. The majority of reads were assigned to bacteria at the domain level with an abundance between 97.78%-98.97%. Archaea, and eukaryote contributed less than 1% abundance (Table 3). In the phylum level, classification an overall dominance of *Proteobacteria*, *Fusobacteria*, *Bacteroidetes*, *Firmicutes*, and unclassified bacteria was observed in the metagenome datasets (Fig. 2a & b). Among these, the most abundant were *Proteobacteria* TP07 (42%) followed by TP06 (41%), TP04 (37%), TP03 (32%), TP01 (26%) TP08 (25%) and TP02 (12%) samples respectively. Whereas, in TP02 the abundance of *Firmicutes* (28%) was higher as compared to other datasets that range from (3 % to 11%). At the class level, taxonomical changes were observed within the stages of CETP metagenome datasets. Within the *Proteobacteria* phylum, Alpha-*Proteobacteria* were found to be most prevalent in TP08 (sludge sample) while Beta-*Proteobacteria* were dominant in TP04 (Fig. 2c). This suggested that the Alpha-*Proteobacteria* class might tolerate well in the CETP treatment process and can be used further for wastewater treatment. At the genus level, the difference between the datasets was also observed, with *Pseudomonas* (7%), *Fusobacterium* (6%), *Bacteroides* (4%), *Leptotrichia* (4%) *Arcobacter* (3.1%), *Methylophage* (3%), *Streptococcus* (2.2%) and *Desulfovomonas* (1.1%) were the most prevalent in the TP01 metagenome dataset. Whereas, in TP02 the dominant genera were *Fusobacterium* (15%) followed by *Clostridium* (12%), *Bacteroides* (11%), *Streptococcus* (8%) and *Desulfovibrio* (4.4 %). The dominance of *Pseudomonas* genus was higher in TP07 (25.1%) dataset followed by TP04 (12.1%) and TP03(8.2%) respectively. However, in TP02 the relative abundance of *Pseudomonas* genus (1.2%) was lower.

Furthermore, the number of shared and unique genera between different metagenomics dataset were represented in the Venn diagram (Fig. 3). The samples were divided into two groups the first group includes CETP inlet (TP01) and sludge samples (TP07 and TP08) and the second group includes inlet going to collection tank (TP01) and the final discharge of the effluent (TP06) (Fig. 3a & b). While, Fig. 3a demonstrated that a total of 69 bacterial genera were common among these metagenomics data sets whereas 126, 336 and 171 bacterial genera are unique in TP01, TP07, and TP08 that resulted in a very diverse community was shared between these metagenomics data sets. Whereas, in the second group 85 genera were found to be common in inlet and outlet of the CETP effluent with 262 bacterial genera were unique to the inlet (TP01) and 136 genera were unique to TP06 i.e. outlet of CETP (Fig. 3b).

At the species level, the CETP Vatva treatment plant was predominantly occupied by *Arcobacter nitofigilis*, *Wolinella succinogenes*, *Desulfovibris vulgaris*, *Desulfovomonas thiophila*, *Clostridium sticklandii*, *Streptococcus suis*, *Thurea* sp, *Enterococcus* sp, *Klebsiella pneumonia*, *Riemerella* sp., *Shewanella* sp., *Alcaligenes* sp. and others. The microbial community was also dominated by different species of *Pseudomonas aeruginosa*, *Pseudomonas filiscindens*, *Pseudomonas fluorescens* and *Pseudomonas stutzeri*. The *Shewanella decolorationis* species was only found in the TP07 whereas, other rare bacterial species across the sampling set were *Acinetobacter* sp., *Aeromonas* sp., *Azospira* sp., *Nitrosomonas* sp., *Rhizobium* sp., and many more.

Functional profiling of the microorganism

The whole metabolic pathways were analysed by using two metabolic databases (KEGG and COG). The results showed that the relative abundance of functional features belonging to COG categories were cellular processes and signalling, information storage and processing, and metabolism (Fig. S2). Metabolism was the most predominant function, ranged from (47.4 to 43.01%) in the CETP samples collected from different sampling points. At sub-level 2, the top six abundant functions were general function prediction, amino acid transport and metabolism, replication recombination and repair, energy production and conservation, translation, ribosomal structure and biogenesis, and cell wall membrane envelope biogenesis in TP01 sample. Whereas, in TP06 dataset top six abundant functions belong to signal transduction, cell cycle control cell division, general function prediction, transcription, cytoskeleton and carbohydrate transport and metabolism, carbohydrate transport and metabolism (Fig. 4a). PCA was performed to investigate the association among the COG functions of the CETP samples. The functional variations were covered in two components PC1 and PC2 with a total variation of 75.8%. In the PCA plot, the samples were clustered into two major groups TP03, TP04, and TP07 and the second major cluster was formed between TP05, TP06, and TP08 while TP01 and TP02 were separated independently (Fig. 4b). The result also supported in the dendrogram and heatmap (Fig. 4c & Fig. S3).

In KEGG annotation, the most abundant function belongs to metabolism, as predicted in COG analysis. KEGG mapping classified the metagenome dataset in metabolism (55.57 to 59.39 %), genetic information processing (18.61 to 21.19%), environmental information processing (11.58 to 16.12%), cellular processes (4.95 to 5.34%), human diseases (0.86 to 1.83%) and organismal systems (0.36 to 0.49%)

(Fig. S4). At KO level 2, carbohydrate metabolism was the most abundant function represented by 19.27 to 22.71% in TP01, TP02, TP03, TP04, TP06, and TP07 samples. Whereas, amino acid metabolism was the most abundant function represented by TP06 (19.93%) and TP05 (19.79%). The other KO categories were energy metabolism (12.9 to 14.83%), metabolism of cofactors and vitamins (10.3 to 11.15%), nucleotide metabolism (9.96 to 11.27%), lipid metabolism (5.41 to 6.58%), metabolism of other amino acids (4.85 to 5.62%), xenobiotics biodegradation and metabolism (2.57 to 4.66%), glycan biosynthesis and metabolism (3.25 to 4.93%), metabolism of terpenoids and polyketides (3.02 to 3.24%) and biosynthesis of other secondary metabolites (1.92 to 2.69%) (Fig 5a). The PCA analysis for the relative abundance of an annotated sequence of functional systems from CETP samples was presented in Fig. 5. In the PCA plot, high variance indicated a statistically significant difference in the abundance of functional profiles among the different sampling points in the CETP samples. The samples were clustered in groups, similarly as analysed in the COG annotations (Fig. 5b). Moreover, results were also supported by the dendrogram and in the heatmap (Fig 5c & Fig. 5S).

KEGG pathway assignment

Pathway prediction was performed based on the KEGG database. Further, each of the enzymes involved in the degradation was filtered and the gene abundance was calculated from each metagenome dataset and was represented in Table S1, S2, S3. It was observed that genes related to the metabolism of the aromatic compound were more abundant in the TP05 metagenome dataset as compared to other sampling datasets.

Benzoate biodegradation pathway

In the metagenomics dataset, a maximum number of genes were annotated from the benzoate degradation pathway. A total of 10,226 genes with 81 KEGG hits were identified and mapped in KEGG mapper. The key enzymes involved in the benzoate degradation pathways via hydroxylation were catechol 1,2-dioxygenase (EC 1.13.11.1), and protocatechuate 3,4-dioxygenase (EC 1.13.11.3), were identified in the CETP metagenomes dataset (Fig. 6; Table S2). These enzymes were involved in the aerobic benzoate degradation using dioxygenase to form catechol and monooxygenase to form protocatechuate. Almost all enzyme involved in the aerobic benzoate degradation was covered by metagenome datasets (Fig. 6). Moreover, the cluster of the *ben* gene (*benK*, *benE*, *benABC*, and *benD*) was also found in the dataset. The *benABCD* and *benD* were involved in converting benzoate to catechol whereas, *benK* and *benE* help in the transportation of benzoate inside the cell and *cat* genes (*catA* and *catE*) is involved in degradation of catechol.

A second mechanism to degrade benzoate is via protocatechuate formation, in which gene *CYP53A1* and *PobA* were involved. The abundance of the *CYP53A1* gene was found only in TP05 and TP06, although the *PobA* gene was present in all samples. After the formation of protocatechuate, *PcaGH* and *LigAB* genes were further involved in ortho and meta cleavage. Moreover, the gene abundance to degrade benzoate via box pathway (aerobic hybrid pathway) was also found in CETP metagenome datasets.

Phenylalanine Metabolism Pathway

Gene mining was done to predict the enzymes for the phenylalanine metabolism pathway in CETP metagenome dataset. The different genes involved in the phenylalanine was found in all the metagenome dataset with their abundance (Table S3). In metagenome dataset, the conserved gene cluster i.e. (*paa* gene cluster) was found in all sampling points which include 14 genes and two catabolic operon *paaXY*. The Fig. 7, represents the different *paa* gene cluster involved in the degradation of phenylalanine. Initially, the catalase-peroxidase enzyme (*Kat G* gene) acts on phenylalanine, and converts it to 2-phenyl acetamide and then phenyl acetamide in the presence of *AmiE* (amidase) enzyme. Further, *paaK* (phenylacetate-CoA ligase EC:6.2.1.30) gene catalyze the reaction by converting phenyl acetyl to phenyl Acetyl-CoA. The abundance of the *paaK* gene was maximum in TP07 (189) followed by TP04 (131), TP03(119), TP01(97), TP02(90), TP05(81), TP08(77) and TP06 (43). The presence of other genes which were involved in the phenylalanine degradation pathway include; *paaG*, *paaJ*, *paaF* and *paaH* genes. The *paaABCD* gene was also found in metagenome dataset and proposed to be responsible for the introduction of oxygen into the aromatic ring of phenylacetyl-CoA. Further, the *paaJ* gene catalysed the reaction of 3-oxo-5,6- di-dehydro suberyl-CoA to Acetyl-CoA or further involved in 2,3-didehydro-adipyl-Co to metabolise via succinyl-CoA, and Acetyl-CoA in the presence of CoA, NAD⁺. Gene *paaF* catalysed the reversible reaction of 2,3-dehydroadipyl-CoA to 3-hydroxyadipyl-CoA and *paaH* gene was involved in last steps of phenylalanine metabolism, which catalysed 3-dehydroadipyl-CoA to 3-oxo-adipyl-CoA, subsequent formation of succinyl-CoA.

Degradation pathways of 1,2-dichloroethane

Enzymes encoded by bacteria enable them to degrade the synthetic chlorinated compound. The presence of a such gene in the effluent provides clear evidence of the utilization of chlorinated compounds in a huge amount by the industries. In the first step of the 1,2-dichloroethane (DCA) degradation, the conversion of 1,2-dichloroethane to 2-chloroethanol occurs in the presence of gene *Dh1A* as shown in Fig. 8. The abundance of this gene was present in all samples except TP02. In alternative pathways for the degradation of DCA alkane-1, monooxygenase play the role in conversion via the oxidation process, because this enzyme has wide substrate specificity and performs both epoxidation and hydroxylation reaction in the industrial effluent. The second step includes a further breakdown of 2-chloroethanol to

chloroacetaldehyde through enzyme methanol dehydrogenase (*mdh1*), this gene was also found in all datasets but absent in the TP02 sample (Table S3). Furthermore, chloroacetaldehyde gets converted to chloroacetate, this reaction was catalyzed by aldehyde dehydrogenase (ALDH) enzymes. In the last step of pathway, the conversion of chloro-acetate to glycolate with the presence of haloacetate dehalogenase and 2-haloacid dehalogenase which further enter into glyoxylate and dicarboxylate metabolism (Fig. 8).

The dominance of the various genes involved in the aromatic compound degradation pathway was also identified in the whole genome sequencing of pure culture isolate. In *Pseudomonas aeruginosa* and *Bacillus licheniformis* genomic data identifies different genes probably involved in the degradation of benzene, phenol, biphenyl and 1-2 dichloromethane pathways (Table S4, S5, S6). The data also revealed that catechol 2,3 dioxygenase genes were present which are involved in the degradation of catechol in the ortho and meta-cleavage pathways respectively (Fig.6). As reported in the supplementary Table S9 and Fig. 6, the strains *P. aeruginosa* genome contains maximum number genes implicated in the degradation of benzoate however in *B. licheniformis* genome contains some of the genes involved in the degradation of the benzoate pathways (Fig. 6,7,8; Table S4). Similarly, both the strains contain the enzyme-coding genes which are involved in the degradation of phenylalanine, and 1,2-dichloroethane.

Discussion

The common effluent treatment plant treats different industrial effluent coming from dye manufacturing industry, chemicals and other industries. During manufacturing process, different dyes are used for the colouring purpose and large amount of effluent and solid waste is generated which is the mixture of several dyes. Before the discharge, the effluent is being treated with a series of process steps which include chemical and biological treatment. The degradation of such effluent required large quantity of organic and dynamics process based on the actual mixture of compounds in the wastewater [13]. However, no specific study that have focused on the changes in the microbial dynamics at different stages of the treatment plant. Therefore, we have collected the water and sludge samples from the different stages to provide insights into the microbial community composition and their diverse function in toxic compound biodegradation. The physicochemical and GC-MS analysis was carried out for the characterization of the waste water environment. During the successive treatment a reduction in Cr, Ni and Zn was observed. This might be due to its bioaccumulation inside the bacterial cells or binding with lipopolysaccharides of the extracellular membranes [14]. The microorganisms belonging to *Pseudomonas*, *Desulfuromonas* has been reported for reducing metals to less or nontoxic metals [15]. However, the Cu concentration was slightly higher in the final outlet as compared to other sampling point that might be due to induction of corrosion with copper plumbing.

GC/MS analysis detected the various compounds which belong to the aromatic compounds, acids and hydrocarbons. The detection of compounds in different sampling points indicates the recalcitrant nature of the compounds as some of them does not degrade completely during the treatment process [3]. The detection of sulphuric acid in all the stages showed that sulphuric acid is being used for maintaining the pH within the wastewater treatment plants. Moreover, the phthalic anhydride was also detected in both in inlet (TP01) and outlet (TP06) of the textile effluent as well as in the sludge samples. Phthalate, such as phthalic acid, diethylhexyl have been listed as priority pollutants by USEPA [16]. The phthalic acid is the aromatic dicarboxylic acid commonly used for the dye manufacturing. The phenolic compound and phthalates are reported as potential endocrine disrupting chemicals. However, the percent peak area was reduced in the outlet of the effluent this indicated the effective degradation of complex compounds. Moreover, most of the organic compounds detected in the inlet of CETP effluent were diminished during the treatment at CETP. The disappearance of most organic compound from CETP may be the utilization of the sole source of carbon, nitrogen from the bacterial community present over there and which play major role in detoxification and degradation of CETP wastewater.

The *Proteobacteria*, *Fusobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* were dominant phyla in all stages of CETP with the exception in TP02 dataset. In TP02 the relative high abundance of *Firmicutes* was observed, which is versatile in degrading a vast array of environmental substrate [17]. They are involved in metabolic pathways responsible for producing volatile fatty acids, which can be utilized by other group of microbes [18]. Compared to previous reports the *Proteobacteria* was detected as the most dominant phylum in various wastewater treatment plants [19]. These bacteria are facultative or obligate anaerobes with a diverse metabolic plasticity and are major contributors for removal of organic and nutrient [20]. Moreover, Chloroflexi phylum abundance was also observed in the last stages of CETP treatment plant i.e TP06 and TP08 sampling points. Bacteria belonging to *Chloroflexi* phyla participate in the degradation of complex organic compounds and polymers. These bacteria decompose dead cells and exopolysaccharides into simple organic molecules includes lactate and ethanol which can be utilized by others species in their metabolism. Thus, these bacteria can be sustained under harsh environmental conditions [21]. The results were comparable with the studies of other wastewater systems [11] [20][21].

A predominance of *Pseudomonas*, *Arcobacter*, *Methylophage*, *Streptococcus*, *Bacteroides* and *Desulfuromonas* was noted for all the samples. The *Pseudomonas* was the most dominant genus among the entire genus. This indicates that these organisms are active under toxic conditions and can survive under oligotrophic environments [22]. Many studies have shown that the different species of *Pseudomonas* are able to degrade the synthetic dyes mainly azo dye in the liquid medium and have been used for the treatment of textile effluent [23]. Moreover, many

sulphonated aromatic compounds are utilized by dye manufacturing industries as primary materials to produce dyes and many of them are released as byproducts in the effluents coming from textile industries [10]. Besides, the textile industry is widely using Na_2SO_4 in dyeing process and sulphuric acid for maintaining the pH within the wastewater treatment plants. Therefore, the presence of sulphur reducing bacteria such as *Desulfuromonas*, *Desulfovibrio*, *Thiobacillus* were found in the CETP plant. In addition, the presence of *Thaurea* genus was detected in all the sample but was dominant in TP04 sample. The bacteria belonging to genus *Thaurea* has been found in waste water systems and are versatile in aromatic compound degradation. Under both aerobic and anaerobic conditions, aromatics such as benzoate degradation was initiated by benzoyl-CoA ligase via CoA thioesterification pathway by *Thaurea aromatic* K172 as reported by [24]. Thus, *Thaurea* is considered as important genus for treatment of all industrial wastewater system due to its vital role in the degradation of aromatic compounds under denitrification [23]. *Pseudomonas* sp., *Klebsiella pneumonia*, *Riemerella* sp., *Shewanella* sp., *Alcaligenes* sp and *Shewanella decolorationis* were detected in metagenomics datasets. *Shewanella* strains were found to be able to decolorize azo dye [25] anthraquinone dye [26]. In this study, different genes encoding various enzymes required for the degradations of aromatic compounds were identified. Different enzymes, such as oxygenases, oxidoreductases, dioxygenases, ligases involved directly and indirectly in bioremediation were detected and mapped for degradation pathways.

Certain bacterial strains showed the abundance in effluent sample and it directly correlated with the metabolic function to degrade the xenobiotic compounds [13]. With this objective the metabolic potential of the dominant microbial community was analysed, which give catabolic potential of the biomass in all samples. KEGG database were used for the skeleton-based classification and pathways maps by the using of KEGG mapper to find the genes which helps in degradation of certain compound in certain pathways. Bacterial community structure and metabolic potential of the microorganism which is presence in wastewater effluent are more diverse. Thus, relative abundance of the functional gene that is related to xenobiotic degradation are higher in textile effluent [27].

Benzoate degradation genes were identified throughout the CETP vatva metagenome datasets. This degradation pathway plays an important role in the degradation of different type of aromatic compounds [28][27][13]. The abundance of functional genes in different stage revealed that the microbial community were metabolized by diverse type of benzoate-containing compounds [27]. The pervious study revealed that the pathways of benzoate degradation in bacteria occurs either through anaerobic or aerobic processes and with the mechanism of degradation via CoA ligation or via hydroxylation [29]. The third hybrid pathway for benzoate degradation occurs by introducing oxygen to activate the aromatic ring for cleavage, this pathway was previously identified as catabolism of benzoate and mineralization of aromatic compounds by [30][31]. The utilization of this pathway as a degradation of xenobiotics may help for microorganism to survive in less or fluctuating oxygen concentration in waste water effluent [29]. The different species of *Rhodopseudomonas*, *Pseudomonas*, and *Acinetobacter* are reported by many researchers as the potential for degradation of benzoate, phenol, and PAH family compounds group of compounds. The CETP metagenome sequences also showed the abundance of different strain of belonging to *Pseudomonas*, and *Acinetobacter* in different stage of treatment and suggested that such organism have catabolic potential to degrade benzoate in natural environments.

Polychlorinated biphenyls (PCBs) are synthetic chlorinated organic compounds, stable in nature and widely used in different industrial application. The genes identified from CETP metagenome lead to confirm effluent contaminated with PCBs and they are very toxic in nature and affect the flora and fauna. Presence of these genes in TP07 and TP08 metagenomes dataset revealed that the presence of toxic compounds in the effluent and the microorganism are utilizing that as an organic source for their metabolism and can act. BphA1 (biphenyl dioxygenases) are the enzyme specific have been reported in several bacterial strain *Paraburkholderia xenovorans* LB400, *Pseudomonas pseudoalcaligenes* KF707, *Rhodococcus jostii* RHA1 [32][33][34] and were also found in CETP metagenome dataset.

Taguchi et al., [35] reported the different strains of *Rhodococcus* encoded enzymes BphA1 which are responsible for the polychlorinated biphenyl degrading and these strain were encoding the enzyme also found in TP04, TP05, TP07 and TP08 metagenome. Second gene cluster *bphDE* and *etbA1AaAbAc* were also involved in the degradation of biphenyl and PCB and these gene cluster was found in *R. jostii* strain which were also present in TP04 and TP05 metagenome. The third cluster only *narB* gene found in metagenome may be other gene missing or not cover during the sequencing. However, strain of *R. jostii* have multiple pathways like *bph* and *etb* gene cluster [36]. Presence of *bph*, *etb* and *nar* gene cluster in CETP metagenome dataset and also identify the bacterial strain according to previous report suggested that the effluent contain chlorinated compounds [36][37].

The end products of the upper degradation pathway are benzoate and 2-hydroxypenta- 2,4-dienoate. They were further catabolized by three different aerobic pathways catechol, protocatechuate, or benzoyl-coAligation [38][39]. The genes *CYP53A1* and *PobA* are actively involved in benzoate degradation via protocatechuate pathway and bacterial strain *Pseudomonas*, *Bordetella*, *Achromobacter*, *Ralstonia* and, *Rhodococcus* are found in metagenome [38][40]. According to Rather et al., [38] benzoate catalysed by *badA* gene (benzoate-CoA ligase) and box genes involved in hybrid pathway are mostly found in facultative microorganism such as *Achromobacter* and *Variovorax* and these microorganisms utilize the benzoate as a carbon and energy source [40]. Many research studies showed that the presence of *Stenotrophomonas* strain are universal member for degradation of aromatics compounds, biphenyl and, PCBs [41] [42]. The abundance of

these strains are present in the CETP dataset. Its presence in the biphenyl-degrading microbial community in metagenome might be due to utilizing the carbon source of the rest of microbial community by cross-feeding on secondary metabolites [41].

Organic substrate and other constituent of aromatic compounds are most abundant environmental pollutants. However, the substantial part of which phenylalanine is metabolized by the specific group of microbial community. The degradation of phenylalanine and its pathway, as well as operon, are well known in culturable bacterial via an anaerobic process. The presence of phenylalanine and phenylacetate catabolic gene cluster in industrial effluent sample indicating that effluent might contain large amount of aromatic compounds [43]. Pathways are still unknown but metabolic gene cluster are recently identified in *Pseudomonas* spp. [44]. In phenylalanine pathway the first enzyme is phenylacetate-CoA ligase which catalyze the reaction and convert phenylacetate to phenylacetyl-CoA via anaerobic process. *In silico* study result of *PaaABCDE* gene cluster revealed that the enzymes are multicomponent oxygenase which acts on CoA-esters and the complex are made up of several monooxygenases enzymes [44].

Conclusions

The present study was undertaken to provide a comprehensive insight into the microbial taxonomical diversity and functional pattern in the CETP effluent across the different stages of the treatment processes. The results showed that most of the microbial diversity changes with the treatment process in which inlet (TP01) of CETP effluent shows higher dominancy of *Proteobacteria* whereas in TP02 the high abundance of *Firmicutes* was observed. The species such as *Pseudomonas* sp., *Klebsiella pneumonia*, *Riemerella* sp., *Shewanella* sp., *Alcaligenes* sp, and *Shewanella decolorationis* were present in the metagenome data set. The abundance of these species indicates the active involvement of microorganisms for dye degradation. The functional profiling of CETP microbiota resulted in the high abundance of the enzyme such as oxygenases, oxidoreductases, dioxygenases, ligases which were involved in the aromatic compound degradation pathway. Among all, the highest abundance of these genes was present in TP05 (4.6%) and the lowest was present in TP02 (2.57%). The profiling of community structure and degradative pathways generated in this study will be useful in designing the bioremediation strategies for the industrial treatment process.

Abbreviations

CETP

common effluent treatment plant; MG-RAST:Metagenomics Rapid Annotation using Subsystem Technology; RDP:Ribosomal Database Project; COGs:Clusters of orthologous Groups; KEGG:Kyoto Encyclopedia of Genes and Genomes; PHAs:Genomes polyhydroxyalkanoates; ICP-MS:Inductively coupled plasma mass spectrometer; QAST:Quality Assessment Tool for Genome Assemblies; RAST:Rapid Annotation Subsystem Technology; ANOVA:one-way analysis of variance; DMRT:Duncan's Multiple Range test; PAST:Paleontological Statistics; PCA:Principal component analysis; TOC:Total organic carbon; OTUs:operational taxonomic unit.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Raw reads of shot gun metagenomics sequencing are deposited to NCBI short read archive (SRA) under the bioproject PRJNA649407.

Competing interests

The authors declare that they have no competing interests

Funding

Department of Science and Technology (DST), Government of Gujarat, Gandhinagar, India.

Author's contributions

PP designed the experiment, did sampling, participated in metagenome sequencing, performed statistical analysis and drafted the manuscript. RK carried out the metagenomic data analysis and participated in drafting the manuscript. DK did sampling, performed metagenome sequencing and data analysis. ZP participated in sampling and performed metagenome sequencing. LP helped to draft the manuscript. CJ conceptualization of study, contributed in interpretation of data, reviewed and corrected the draft of manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful to the Secretary, Department of Science and Technology (DST), Government of Gujarat, Gandhinagar, Gujarat, India. The author would like to acknowledge the Chairman, GESCSL, Phase-II, GIDC Estate, Vatva, Ahmedabad, Gujarat for extending support in sample collection. Authors would like to acknowledge Dr. Ramesh Pandit and Dr. Neelam Nathani, for critically reviewing and providing essential inputs in the manuscript, Dr. Apurvashin Puvar for providing additional support for Data analysis.

Authors details

Gujarat Biotechnology Research Centre (GBRC), Department of Science and Technology (DST), 6th Floor, MS Building, Gandhinagar, Gujarat, India – 382011

References

1. Paul D, Pandey G, Pandey J, Jain RK. Accessing microbial diversity for bioremediation and environmental restoration. *Trends Biotechnol.* 2005;23:135–42.
2. Moosvi S, Madamwar D. An integrated process for the treatment of CETP wastewater using coagulation, anaerobic and aerobic process. *Bioresour Technol.* 2007;98:3384–92.
3. Chandra R, Bharagava RN, Kapley A, Purohit HJ. Bacterial diversity, organic pollutants and their metabolites in two aeration lagoons of common effluent treatment plant (CETP) during the degradation and detoxification of tannery wastewater. *Bioresour Technol.* 2011;102:2333–41.
4. More RP, Mitra S, Raju SC, Kapley A, Purohit HJ. Mining and assessment of catabolic pathways in the metagenome of a common effluent treatment plant to induce the degradative capacity of biomass. *Bioresour Technol.* 2014;153:137–46.
5. Ju F, Beck K, Yin X, Maccagnan A, McArdeell CS, Singer HP, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* 2019;13:346–60.
6. Cheema S, Bassas-Galia M, Sarma PM, Lal B, Arias S. Exploiting metagenomic diversity for novel polyhydroxyalkanoate synthases: Production of a terpolymer poly(3-hydroxybutyrate-co-3-hydroxyhexanoate-co-3-hydroxyoctanoate) with a recombinant *Pseudomonas putida* strain. *Bioresour Technol.* 2012;103:322–8.
7. Shanks OC, Newton RJ, Kelty CA, Huse SM, Sogin ML, McLellan SL. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. *Appl Environ Microbiol.* 2013;79:2906–13.
8. Riquelme Breazeal M V, Novak JT, Vikesland PJ, Pruden A. Effect of wastewater colloids on membrane removal of antibiotic resistance genes. *Water Res.* 2013;47:130–40.
9. Das S, Bora SS, Yadav RNS, Barooah M. A metagenomic approach to decipher the indigenous microbial communities of arsenic contaminated groundwater of Assam. *Genomics Data.* 2017;12:89–96.
10. Schwitzguébel J-P, Aubert S, Grosse W, Laternus F. Sulphonated aromatic pollutants. *Environ Sci Pollut Res.* 2002;9:62–72.
11. Hu M, Wang X, Wen X, Xia Y. Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresour Technol.* 2012;117:72–9.
12. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 2017;45:W180–8.
13. Jadeja NB, Purohit HJ, Kapley A. Decoding microbial community intelligence through metagenomics for efficient wastewater treatment. *Funct Integr Genomics.* 2019;19:839–51.
14. Srinath T, Verma T, Ramteke PW, Garg SK. Chromium (VI) biosorption and bioaccumulation by chromate resistant bacteria. *Chemosphere.* 2002;48:427–35.
15. Igiri BE, Okoduwa SIR, Idoko GO, Akabuogu EP, Adeyi AO, Ejiogu IK. Toxicity and Bioremediation of Heavy Metals Contaminated Ecosystem from Tannery Wastewater: A Review. *J Toxicol.* 2018;2018.
16. He L, Gielen G, Bolan NS, Zhang X, Qin H, Huang H, et al. Contamination and remediation of phthalic acid esters in agricultural soils in China: a review. *Agron Sustain Dev.* 2015;35:519–34.

17. Liu C, Li H, Zhang Y, Si D, Chen Q. Evolution of microbial community along with increasing solid concentration during high-solids anaerobic digestion of sewage sludge. *Bioresour Technol* [Internet]. The Author(s); 2016;216:87–94. Available from: <http://dx.doi.org/10.1016/j.biortech.2016.05.048>
18. Nascimento AL, Souza AJ, Andrade PAM, Andreote FD, Coscione AR, Oliveira FC, et al. Sewage sludge microbial structures and relations to their sources, treatments, and chemical attributes. *Front Microbiol*. 2018;9:1–11.
19. Cydzik-Kwiatkowska A, Zielińska M. Bacterial communities in full-scale wastewater treatment systems. *World J Microbiol Biotechnol*. 2016;32:1–8.
20. Xia X, Zhang S, Li S, Zhang L, Wang G, Zhang L, et al. The cycle of nitrogen in river systems: Sources, transformation, and flux. *Environ Sci Process Impacts*. 2018;20:863–91.
21. Samson R, Shah M, Yadav R, Sarode P, Rajput V, Dastager SG, et al. Metagenomic insights to understand transient influence of Yamuna River on taxonomic and functional aspects of bacterial and archaeal communities of River Ganges. *Sci Total Environ*. 2019;674:288–99.
22. Atashgahi S, Aydin R, Dimitrov MR, Sipkema D, Hamonts K, Lahti L, et al. Impact of a wastewater treatment plant on microbial community composition and function in a hyporheic zone of a eutrophic river. *Sci Rep*. 2015;5:1–13.
23. Zhang B, Xu X, Zhu L. Structure and function of the microbial consortia of activated sludge in typical municipal wastewater treatment plants in winter. *Sci Rep*. 2017;7:1–11.
24. Schu K, Gescher J, Feil U, Paul M, Jahn M, Scha H, et al. Benzoate-Coenzyme A Ligase from. *Society*. 2003;185:4920–9.
25. Wu J, Kim KS, Sung NC, Kim CH, Lee YC. Isolation and characterization of *Shewanella oneidensis* WL-7 capable of decolorizing azo dye Reactive Black 5. *J Gen Appl Microbiol*. 2009;55:51–5.
26. Xu M, Guo J, Cen Y, Zhong X, Cao W, Sun G. *Shewanella decolorationis* sp. nov., a dye-decolorizing bacterium isolated from activated sludge of a waste-water treatment plant. *Int J Syst Evol Microbiol*. 2005;55:363–8.
27. Balcom IN, Driscoll H, Vincent J, Leduc M. Metagenomic analysis of an ecological wastewater treatment plant's microbial communities and their potential to metabolize pharmaceuticals [version 1; referees: 1 approved, 1 approved with reservations]. *F1000Research*. 2016;5.
28. Díaz E. Bacterial degradation of aromatic pollutants: A paradigm of metabolic versatility. *Int Microbiol*. 2004;7:173–80.
29. Sierra-García IN, Alvarez JC, De Vasconcellos SP, De Souza AP, Dos Santos Neto EV, De Oliveira VM. New hydrocarbon degradation pathways in the microbial metagenome from brazilian petroleum reservoirs. *PLoS One*. 2014;9.
30. Bains J, Leon R, Temke KG, Boulanger MJ. Elucidating the reaction mechanism of the benzoate oxidation pathway encoded aldehyde dehydrogenase from *Burkholderia xenovorans* LB400. *Protein Sci*. 2011;20:1048–59.
31. Valderrama JA, Durante-Rodríguez G, Blázquez B, García JL, Carmona M, Díaz E. Bacterial degradation of benzoate: Cross-regulation between aerobic and anaerobic pathways. *J Biol Chem*. 2012;287:10494–508.
32. Seeger M, Timmis KN, Hofer B. Conversion of chlorobiphenyls into phenylhexadienoates and benzoates by the enzymes of the upper pathway for polychlorobiphenyl degradation encoded by the *bph* locus of *Pseudomonas* sp. strain LB400. *Appl Environ Microbiol*. 1995;61:2654–8.
33. Seto M, Kimbara K, Shimura M, Hatta T, Fukuda M, Yano K. A novel transformation of polychlorinated biphenyls by *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol*. 1995;61:3353–8.
34. Tairazf K, Hiroseji J, Hayashidall S, Furukawasll K. Analysis of. 1992;267:4844–53.
35. Taguchi K, Motoyama M, Iida T, Kudo T. Polychlorinated biphenyl/biphenyl degrading gene clusters in *Rhodococcus* sp. K37, HA99, and TA431 are different from well-known *bph* gene clusters of *Rhodococci*. *Biosci Biotechnol Biochem*. 2007;71:1136–44.
36. Iwasaki T, Miyauchi K, Masai E, Fukuda M. Multiple-subunit genes of the aromatic-ring-hydroxylating dioxygenase play an active role in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol*. 2006;72:5396–402.
37. Yang X, Liu X, Song L, Xie F, Zhang G, Qian S. Characterization and functional analysis of a novel gene cluster involved in biphenyl degradation in *Rhodococcus* sp. strain R04. *J Appl Microbiol*. 2007;103:2214–24.
38. Rather LJ, Knapp B, Haehnel W, Fuchs G. Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *J Biol Chem*. 2010;285:20615–24.
39. Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds- From one strategy to four. *Nat Rev Microbiol*. 2011;9:803–16.
40. Garrido-Sanz D, Manzano J, Martín M, Redondo-Nieto M, Rivilla R. Metagenomic analysis of a biphenyl-degrading soil bacterial consortium reveals the metabolic roles of specific populations. *Front Microbiol*. 2018;9:1–13.
41. Wald J, Hroudova M, Jansa J, Vrchotova B, Macek T, Uhlik O. *Pseudomonads* rule degradation of polyaromatic hydrocarbons in aerated sediment. *Front Microbiol*. 2015;6.
42. Uhlik O, Musilova L, Ridl J, Hroudova M, Vlcek C, Koubek J, et al. Plant secondary metabolite-induced shifts in bacterial community structure and degradative ability in contaminated soil. *Appl Microbiol Biotechnol*. 2013;97:9245–56.

43. Devpura N, Jain K, Patel A, Joshi CG, Madamwar D. Metabolic potential and taxonomic assessment of bacterial community of an environment to chronic industrial discharge. *Int Biodeterior Biodegrad.* 2017;123:216–27.
44. Teufel R, Mascaraque V, Ismail W, Voss M, Perera J, Eisenreich W, et al. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci U S A.* 2010;107:14390–5.

Tables

Table 1

Physicochemical parameters of the different sampling points of common effluent treatment plant (CETP) collected for metagenomics analysis.

Physicochemical parameters	Sampling points (CETP)							
	TP01	TP02	TP03	TP04	TP05	TP06	TP07	TP08
Heavy metal (mgkg ⁻¹)								
Chromium	0.264 ± 0.001 ^a	1.09 ± 0.01 ^b	0.247 ± 0.001 ^c	0.281 ± 0.002 ^d	0.1975 ± 0.001 ^e	0.195 ± 0.001 ^e	0.379 ± 0.001 ^f	0.171 ± 0.001 ^f
Zinc	0.0564 ± 0.002 ^a	0.0716 ± 0.001 ^b	0.076 ± 0.001 ^c	0.077 ± 0.002 ^c	0.0718 ± 0.002 ^b	0.046 ± 0.002 ^d	0.041 ± 0.002 ^e	0.063 ± 0.002 ^f
Copper	0.2374 ± 0.002 ^a	0.9285 ± 0.001 ^b	0.315 ± 0.001 ^c	0.285 ± 0.002 ^d	0.843 ± 0.002 ^e	1.539 ± 0.002 ^f	0.085 ± 0.002 ^g	1.287 ± 0.001 ^h
Nickel	0.1016 ± 0.002 ^a	0.068 ± 0.002 ^b	0.107 ± 0.002 ^c	0.106 ± 0.002 ^c	0.084 ± 0.001 ^d	0.087 ± 0.002 ^d	0.056 ± 0.002 ^e	0.078 ± 0.001 ^f
Iron	1.4205 ± 0.001 ^a	0.3731 ± 0.002 ^b	1.453 ± 0.003 ^c	1.920 ± 0.01 ^d	1.401 ± 0.002 ^e	0.324 ± 0.002 ^f	0.517 ± 0.001 ^g	0.442 ± 0.003 ^h
TOC (ppm)	24.69 ± 0.01 ^a	178.8 ± 0.20 ^b	183.65 ± 0.06 ^c	179.45 ± 0.025 ^d	148.9 ± 0.10 ^e	110.5 ± 0.10 ^f	414.35 ± 0.015 ^g	212.4 ± 0.152 ^h
Means with different superscripts in the same column are significantly different at P < 0.05 (ANOVA, Duncan's Multiple Range Test, SPSS version 25.0) (TP01: CETP Collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).								

Table 2
Summary of metagenome datasets characteristic and annotation.

Metagenome	TP01	TP02	TP03	TP04	TP05	TP06	TP07	TP08
Number of reads	363,533,019	868,143,304	523,590,389	465,247,586	837,421,017	742,336,552	934,630,436	712,002,016
Sequences bases	20,19,140	28,89,778	18,20,675	15,60,732	31,96,042	26,07,051	37,03,776	32,03,598
Average read Length (bases)	180 ± 64 bp	300 ± 84 bp	288 ± 85 bp	298 ± 81 bp	262 ± 87 bp	285 ± 80 bp	252 ± 99 bp	222 ± 89 bp
Average GC distribution (%)	44 ± 12%	43 ± 10%	45 ± 11%	46 ± 11%	50 ± 11%	45 ± 9%	49 ± 12%	51 ± 11%
QC passed reads	357,377,813	853,671,600	306,231,019	457,829,560	825,483,132	735,745,628	919,620,613	701,359,377
Mean Sequence Length	180 ± 63 bp	301 ± 82 bp	188 ± 84 bp	298 ± 80 bp	262 ± 86 bp	285 ± 79 bp	253 ± 98 bp	222 ± 88 bp
Predicted Protein Features	13,09,352	1537288	10,06,523	12,15,848	19,05,843	15,01,822	22,57,602	18,40,316
Predicted rRNA Features	5,543	7343	5,124	7,280	6,027	4,123	9,852	4,987
Identified Protein Features	6,56,582	813718	4,94,626	6,70,144	7,23,462	4,12,742	11,81,699	6,73,076
Identified rRNA Features	3,240	3695	3,003	4,316	2,848	2,010	5,429	2,447
<i>(TP01: CETP Collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal)</i>								

Table 3 Summary of diversity indices for taxonomic assignment of bacterial diversity of CETP, Vatva

Diversity Indices	TP01	TP02	TP03	TP04	TP05	TP06	TP07	TP08
Taxa_S	347	379	357	437	300	221	592	305
Individuals	4558	7370	4504	5322	2545	1532	7880	2205
Dominance (D)	0.22	0.06	0.16	0.15	0.09	0.09	0.09	0.12
Simpson's Index of Diversity (1-D)	0.77	0.93	0.83	0.84	0.90	0.90	0.90	0.87
Shannon's Diversity Index (H)	3.05	3.95	3.32	3.53	3.80	3.45	3.89	3.68
Evenness_e^H/S	0.06	0.13	0.07	0.07	0.14	0.14	0.08	0.13
Fisher_alpha Chao1 (1/D)	742.80	570.80	603.1.	802.20	547.20	529.8	1018	541.5

(TP01: CETP Collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).

Table 4 Domain distribution of different sampling points of CETP, Vatva metagenomics datasets

	TP01	TP02	TP03	TP04	TP05	TP06	TP07	TP08
Bacteria	98.82 %	98.97%	98.00%	98.59%	98.43%	98.89%	98.07%	97.78%
Archaea	0.02%	0.01%	0.02%	0.02%	0.04%	0.07%	0.01%	0.05%
Eukaryota	0.00%	0.00%	0.02%	0.00%	0.51%	0.26%	0.10%	0.54%
unclassified sequences	1.16%	1.03%	1.98%	1.41%	1.02%	0.78%	1.79%	1.68%

(TP01: CETP Collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).

Figures

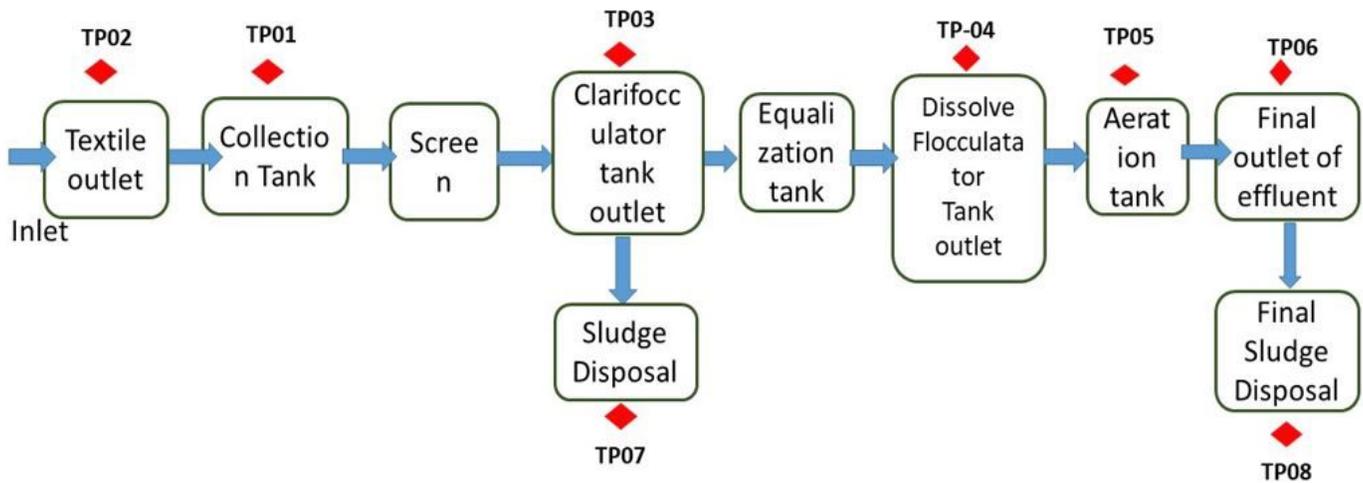


Figure 1

Sampling points from CETP, Vatva (TP01: CETP collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air Flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).

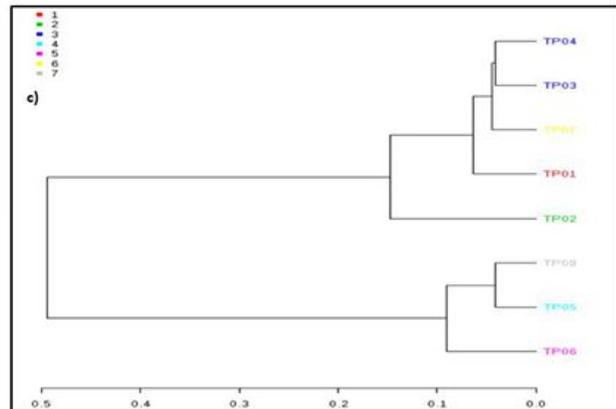
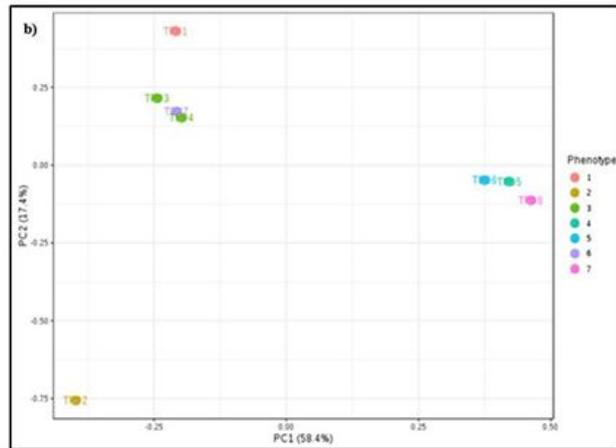
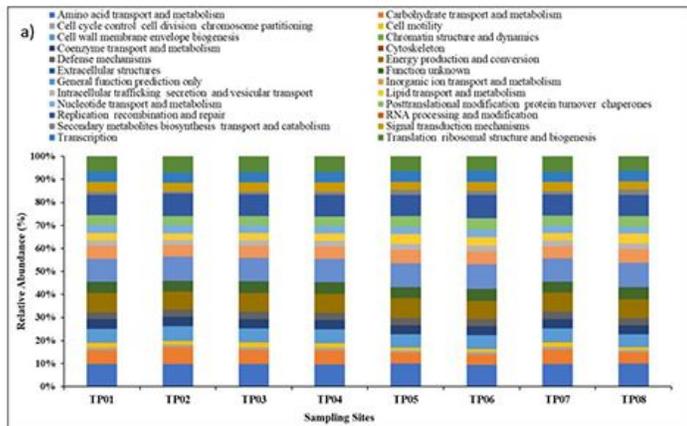


Figure 4

a) Functional profiling from eight different sampling points in CETP vatva based on COG classification at level 2 a) relative gene abundance (%) b) scatter plot of principal component analysis (PCA) depicting variance of COG categories detected in different metagenome dataset c) clustering result shown as dendrogram (distance measure using bray and clustering algorithm) (TP01: CETP collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air Flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).

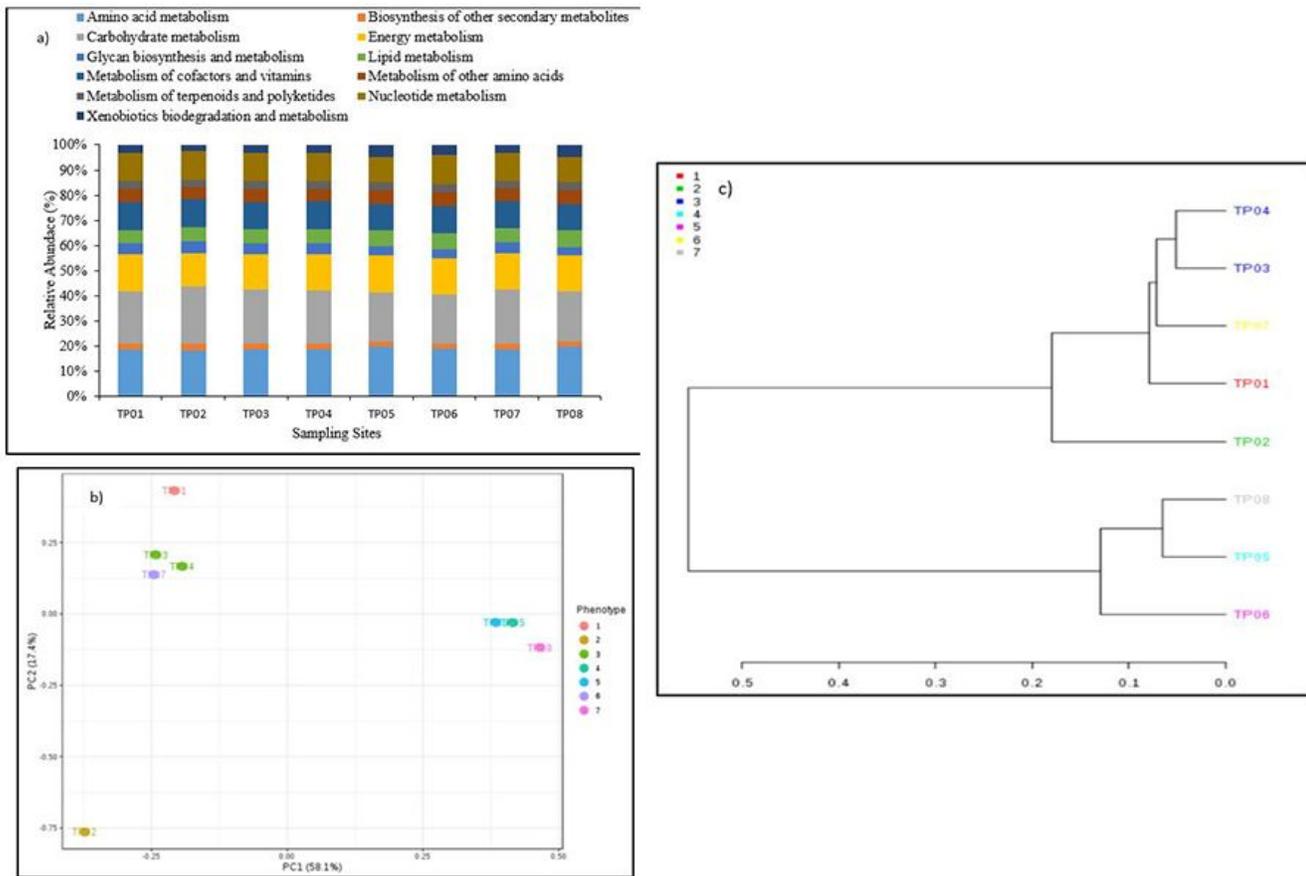


Figure 5

KO based mapping of metagenomics derived ORFa at level 2 a) relative gene abundance (%) b) scatter plot of principal component analysis (PCA) depicting variance of KEGG functions detected in different metagenome dataset c) clustering result shown as dendrogram (distance measure using bray and clustering algorithm) (TP01: CETP collection tank; TP02: Textile industry effluent; TP03: Clariflocculator tank outlet; TP04: Dissolved air Flotation tank outlet; TP05: Aeration tank; TP06: Final outlet of effluent; TP07: Sludge disposal; TP08: Final sludge disposal).

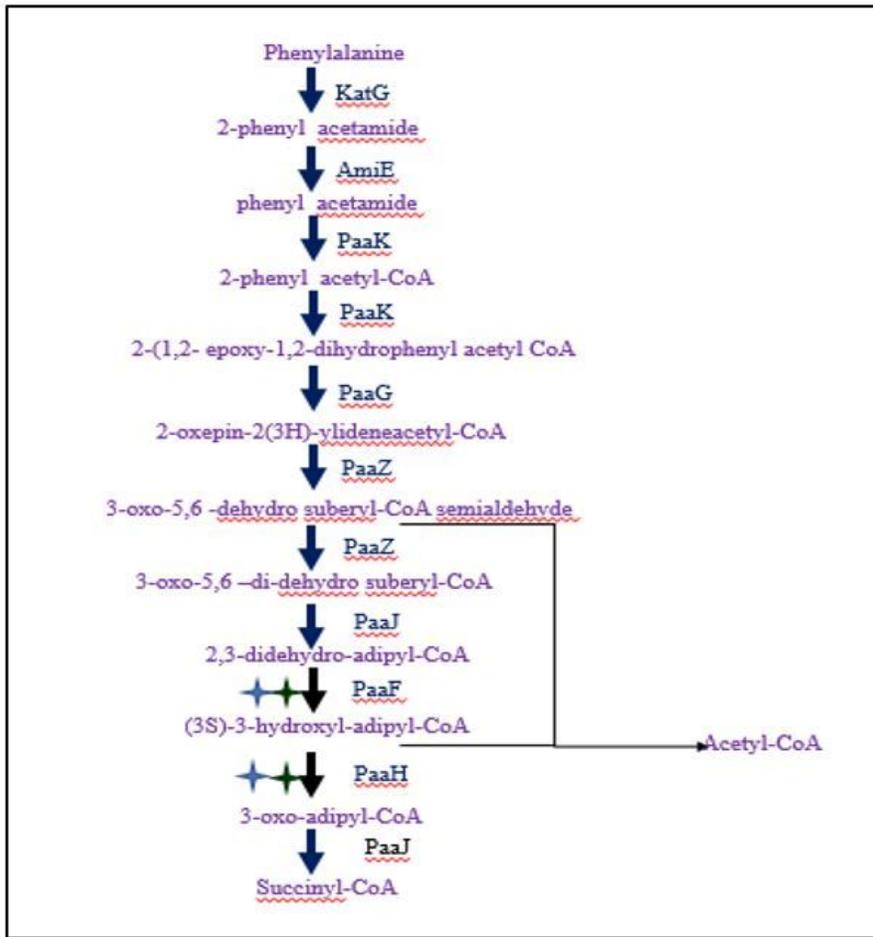


Figure 7

Predicted pathway for phenylalanine degradation identified from CETP, Vatva metagenome dataset based on KEGG pathway database. The star marked symbol represents the enzymes identified from whole genome of isolate *Pseudomonas aeruginosa* () and *Bacillus licheniformis* ().

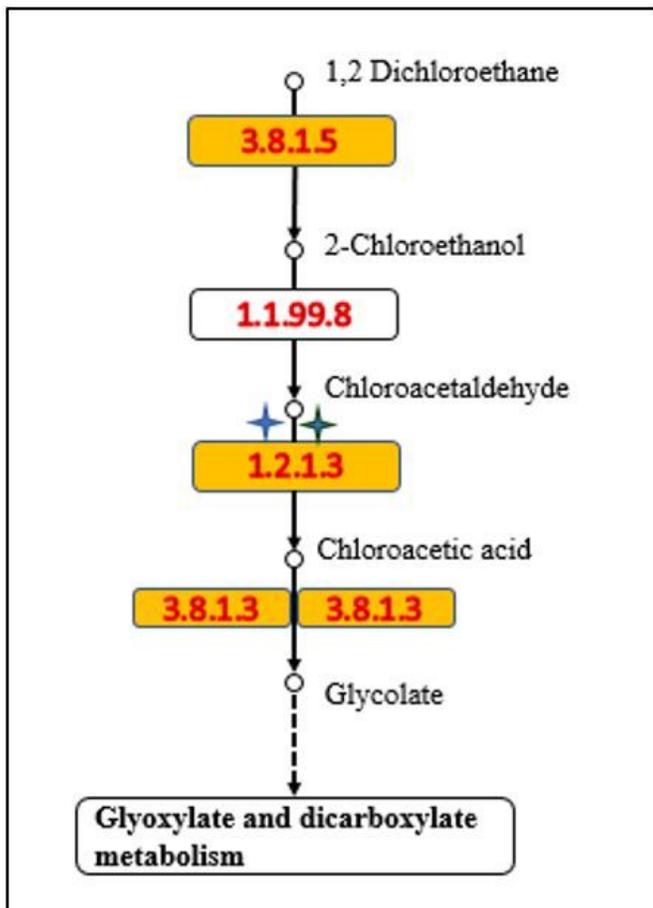


Figure 8

Predicted pathway for 1,2 dichloroethane. Different participating enzymes are represented by enzyme commission number rectangles; yellow fillings indicate enzyme-coding genes identified from CETP metagenome dataset based on KEGG database. The star marked symbol represents the enzymes identified from whole genome of isolate *Pseudomonas aeruginosa* () and *Bacillus licheniformis* ().

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [suplementryData.docx](#)