Metabolic potential and taxonomic assessment of bacterial community of an environment to chronic industrial discharge

Neelam Devpura a, Kunal Jain a, Anand Patel b, C.G. Joshi b, Datta Madamwar a, *

a Environmental Genomics and Proteomics Lab, UGC Centre of Advanced Study, Post Graduate Department of Biosciences, Satellite Campus, Vadtal Road, Sardar Patel University, Bakrol, 388 315 Gujarat, India
b Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Anand, 388 001 Gujarat, India

Article info
Article history:
Received 20 October 2016
Received in revised form
14 June 2017
Accepted 15 June 2017
Available online 13 July 2017

Keywords:
Industrial pollution
Xenobiosis
Aromatic compounds
Stress response genes

Abstract
Due to the anthropogenic activities several pristine ecosystems have being altered. Different bioremediation studies have emphasized the synergistic role of diverse microbial (bacterial) community in metabolism of xenobiotic compounds. Thus, for better understanding of native bacterial community, catabolic and phylogenetic diversity of polluted soils of Sachin Industrial Estate, Sachin, Gujarat, India was studied using high throughput sequencing technology. In this study we have annotated genes such as mono- and dioxygenase, dehydrogenase, metal reductases, etc. reported for degradation of substituted and non-substituted aromatic, higher chain hydrocarbons and aliphatic compounds, resistance and detoxification of heavy metals. Different genes involved in protection from ROS and osmotic stress were also detected. The bacterial population were predominated by Actinomycetales, Rhizobiales and Enterobacteriales of phyla Actinobacteria and Proteobacteria respectively. The functional diversity analysis also indicated the active role of bacterial species belonging to above taxa for metabolism of aromatic compounds and heavy metals. Statistical analysis revealed that bacterial community are diverse and equally rich at polluted sites. Therefore, the study has revealed that environment polluted by industrial activity, contains diverse bacterial community possessing genes for degradation of xenobiotic compounds.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction
Anthropogenic activities over the past centuries have released the compounds of xenobiotic origin (which are seldomly found in the nature) in various quantities in open environment (Ojo, 2007). Some of these compound are simpler and being consumed early without affecting the microbial community at the disposal sites, while some xenobiotics are highly recalcitrant, which persist longer in the environment (Stenuit et al., 2008). Xenobiotics in such cases exert selective pressure on microbial communities, which gradually eliminate the sensitive species allowing the perpetuation of adapted communities (Nojiri et al., 2004). Such resemblance in biochemical features may be attributed to the similarity in gene plasticity across the phyla (Top et al., 1995; Herrick et al., 1997; Vallaeys et al., 1999; Poelarends et al., 2000; Senthilo et al., 2000; Springael et al., 2001). Therefore, knowing the phylogenetic diversity of polluted environment would readily aid in understanding the type of community involved in sustaining the perturbed ecology which inturn helps in the bioremediation process.

Determining the metabolic processes performed by the microbes in contaminated niches is imperative for understanding, manipulating and restoring the contaminated ecosystems. However, our knowledge about the taxonomic and phylogenetic relationship amongst the prokaryotes seamlessly expanded across the biomes, the change in functional capabilities of the community is still overlooked (Fierer et al., 2012). Under the selective pressure of xenobiotic compounds, microbial communities have evolved the necessary novel enzymatic process capable of modifying or degrading the recalcitrant compounds (Lorenzo, 2001). At the same time, with the increase in substrate range of existing enzymes new degradative metabolic pathways have been evolved through

* Corresponding author.
E-mail addresses: ndneelam13@gmail.com (N. Devpura), datta_madamwar@yahoo.com (D. Madamwar).

http://dx.doi.org/10.1016/j.ibiod.2017.06.011
0964-8305 © 2017 Elsevier Ltd. All rights reserved.
natural selection (Galvao et al., 2005). The polluted habitats, therefore, have become the largest reservoir of genes that have co-evolved with the structural complexity of xenobiotics (Galvao et al., 2005). Moreover, evolution of degradative metabolic pathways for recalcitrant substrates is enormously significant for remediation of contaminated environment and they would be equally essential as the sources of novel catalytic activities (Timmis and Pieper, 1999; Schmid et al., 2001).

By knowing the taxonomic identification, we might anticipate an overall correlation between the phylogeny and functional elements of the communities in the perturbed habitat, however this may provide us limited assumptions. Because, distinct taxa share specific functional capabilities, while closely related taxa may have different physiologies and environmental tolerance (Fierer et al., 2012). As a result, we cannot completely rely on perception of taxonomic or phylogenetic structure of microbial communities to predict the functional capabilities or functional diversity of these communities (Green et al., 2008). Therefore, studying the genes, gene clusters or ORFs along with the phylogenetic diversity of the contaminated environment will improve our understanding about the functional systems involved in xenobiotic metabolism.

In a light of above perceptions, a study has been designed to establish the functional competence of the bacterial community and their phylogenetic/taxonomic relationship of polluted sites of Sachin Industrial Estate, Sachin, Surat, Gujarat, India.

2. Materials and methods

2.1. Site description and sampling

Sachin Industrial Estate at Sachin, Gujarat, India, is the second largest industrial hub (in terms of the area covered) in the Asian continent, with diverse range of industrial units such as dyes and dye intermediates manufacturing, printing, weaving and fabric processing, wooden and lamination units, textile machineries, etc. The incessant manufacturing in the estate since 1980 usually releases the liquid waste industrial effluents into the nearby constructed open canal (Haq and Chakrabarti, 2000). The soil characteristics are as mentioned in Table S1.

Subsurface soil samples (5–7 cm below) were collected from three different sites in the range of 200 m² area (21.08°N 72.88°E). The sites were selected where effluents from different industries manufacturing wide range of products are constantly being released. From each sites three replicates (to avoid the biasness and to gain the comprehensive understanding of microbial life prevailing at polluted environment) were sampled randomly and transported to the lab under below ambient conditions (4–8 °C). The soils were homogenized, plant debris, roots and other coarse particles were removed and metagenomic DNA was extracted within 24 h of sampling.

2.2. Metagenomic DNA extraction and sequencing

The metagenomic DNA from the each replicates of polluted soil samples were independently extracted by the method described in Desai and Madamwar (2007) with following modifications. The extracted metagenomic DNA was purified by silica based gel extraction kit (Merck Biosciences, Bangalore, India) as per manufacturer’s instruction. The present study was aimed at understanding the functional potential of the microbial community (especially bacterial genes) from environment polluted by industrial activities and their taxonomic identification and phylogenetic affiliation, but it does not focused on the diversity differences of microbial community of individual locations. Therefore, metagenomic DNA extracted from each replicate were pooled in equi-millimolar ratio to make a composite DNA representing the polluted environment of Sachin Industrial Estate, Sachin. The composite sample was prepared for shotgun sequencing on Ion Torrent Platform using 316 chip, with 300 bp chemistry as per manufacturer instructions.

2.3. Post sequencing analysis

Post-sequencing filtered sequences were uploaded on MG-RAST server for functional and taxonomic annotations. The functional annotations of the reads were performed using Subsystem (SEED) database of the MG-RAST server. The taxonomic identification of the reads was performed using three databases on the MG-RAST server, (1) M5NR, (2) IMG and (3) LCA. Each reads for their functional attributes and phylogenetic affiliation were annotated at a sequence similarity of >60% confidence level, having e-value cutoff of $1 \times 10^{-5}$ and minimum nucleotide alignment length of >15 nucleotides (Hinsu et al., 2017; Jadeja et al., 2014; Pruesse et al., 2007; DeSantis et al., 2006; Cole et al., 2003). The alpha diversity was estimated using PAST3 software with default parameters (Hammer et al., 2001).

3. Results

3.1. Microbial functional pattern

The obtained results and gene annotation study revealed that genes encoding for different enzymes can be clustered together and broadly categorized into two groups, housekeeping genes: for those enzymes required for central metabolism of all microbes and for maintaining cell viability, while second category: potentially essential genes encoding the enzymes catalyzing the niche specific functions responding and adapting the environmental change like genes for xenobiotic compound degradation and their metabolism. The first category of functions generally includes the genes required for central metabolism, energy conservation, cell division, cellular processes and signalling, amino acid, lipid, protein, DNA, RNA metabolism, membrane transports and many more. Another category of functional genes are expressed under certain environmental conditions like degradation of xenobiocmpounds (benzene, chloroaromatic, toluene, naphthalene, etc.) heavy metal resistance, drug alteration, stress activated as well as response genes and others.

3.2. Genes and functions associated with organic contaminant degradation

3.2.1. Aromatics degradation genes

Aromatic compounds are usually one of the major pollutants at the sites contaminated with industrial activities. Their metabolism by bacteria (and other microbes) in various environment can broadly be categorized into peripheral pathways for the catabolism of aromatic compounds, metabolism of central aromatic intermediates, anaerobic degradation of aromatic compounds. The functional capabilities for metabolism of different aromatic compounds of native microbial community were characterized into fifteen distinct pathways and clusters as observed from Fig. 1. The figure revealed the abundance of genes for degradation of compound of benzoate origin under anaerobic conditions (i.e. anaerobic benzoate metabolism), followed by abundance for catechol branch of β-ketoacid pathway and phenylacetyl-CoA catabolic pathway. Several different genes encoding for various enzymes, catalyzing the conversion of aromatic compounds (i.e. degradation of aromatic compounds) into simpler, non-cyclic lower molecular weight compounds were annotated from the polluted
metagenome. The gene annotation results indicated that the different substituted and non-substituted phenolic compounds (like 4-hydroxyphenylacetate) at the polluted sites of Sachin, were degraded by enzyme 4-hydroxyphenylacetate-3-monoxygenase (EC 1.14.13.3), which is also involved in metabolism of aromatic biogenic amine and 4-hydroxyphenylacetate degradation. Genes coding for 4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27), required in catabolism of phenylalanine, aromatic amino acid and other aromatic intermediates were detected. They are also involved in tryosine metabolism.

Few of the benzoate degradation genes encoding for benzoate-CoA ligase (EC 6.2.1.25), essentially required for aminobenzoate degradation, aerobic and anaerobic benzoate and other similar aromatic compound degradation, while benzoyl-CoA reductase (EC 1.3.99.15) also involved in benzoate and benzyl-CoA degradation (anaerobic), protocatechate 4.5-dioxygenase (EC 1.13.11.8) catalyzing the degradation of 3,4- dichlorobenzoate, 3-benzoate, aninobenzoate, and polycyclic aromatic hydrocarbon degradation were annotated (Table S2/Aromatic Compounds). The study further revealed the detection of genes for benzaldehyde dehydrogenase oxidoreductase (EC 1.2.1.28) required for conversion of m-xylene degradation to m-toluate, p-xylene degradation to p-toluate, tolunte to benzoate, degradation of 3-chlorotoluene and amimo-benzoate. The annotation also indicated the presence of genes for biphenyl-2,3-diol-1,2-dioxygenase (EC 1.13.11.39), an enzyme involved in 2-hydroxybiphenyl, carbazole, chlorobenzene and diphenyl ethers degradation. It also involves in degradation cascade of biphenyl and poly chlorinated biphenyls. Other genes (important in aromatic metabolism), like, catechol 1,2-dioxygenase (EC 1.13.11.1), gentisate 1,2-dioxygenase (EC 1.13.11.4), homogentisate 1,2-dioxygenase (EC 1.13.11.5) involved in 3-chlorocatecho, toluene, fluoroobenzolate, gentisate, 2,5-xenol and 3,5-xenol, salicylate and styrene degradation respectively were observed.

Genes for metabolism of central aromatic intermediates were mainly identified from Rhizobiales, Burkholderiales, Actinomycetales, Chloroflexales, Rhodobacterales, Pseudomonadales and other group of bacteria. The key degraders were Methylobacterium sp., Rhodopseudomonas sp. Bradyrhizobium sp., Bordetella sp.,Ralstonia sp., Rhodoferax sp., Roseiflexus sp., Paracoccus sp., Mycobacterium sp., Streptomycetes sp., Pseudomonas sp. The genes for peripheral pathways for the catabolism of aromatic compounds were detected from Actinomycetales, Burkholderiales, Bacillales, Myxococcales, Sphingomonadales Pseudomonadales, Enterobacterales, Rhodocyclales and other orders of the bacterial phylony. Essentially, Mycobacterium spp., Streptomycetes spp., Rhodococcus sp., Acidovorax sp., Bordetella sp., Burkholderia sp., Anaeromyxobacter sp., Methylobacterium sp., Rhodopseudomonas sp., Sphingomonas sp., Escherichia sp., Aromatoleum sp., Azorarcus sp., Silicibacter sp., were detected for harbouring the genes for degradation of xenobiotic aromatic compounds through peripheral pathways.

3.2.2. Heavy metal resistance genes

Heavy metals are also major components of sites contaminated with industrial activities and majority of the heavy metals possess toxicological properties. The functional capabilities for metabolism of resistance towards heavy metals, antibiotics and drugs of native microbial community were characterized into fourteen distinct pathways and clusters as observed from Fig. 2. From the figure it can be observed that genes for cobalt-zinc-cadmium resistance were highly abundant followed by genes for multidrug resistance efflux pumps (complete pathway) and resistance towards fluoroquinolones. Fluoroquinolones and quinolones are both naturally derived and synthetic broad-spectrum antibiotics (Heeb et al., 2011).

For arsenic and chromium resistance, arsenic efflux pump protein, arsenical pump-driving ATPase (EC 3.6.3.16), arsenical resistance operon repressor were annotated, chromate transport protein ChrA respectively were annotated (Table S2/Heavy Metals and Antibiotics). Both, organomercuric [Hg(II)]^+ and Hg(II)^+ are observed for their highly toxic nature to living organisms (Hong et al., 2010). Gene involved in demethylation of reduced mercury...
(an important step for mercury recycling) (Parks et al., 2009), organomercurial lyase (EC 4.99.1.2) degrades phenylmercury acetate, was annotated from the metagenome. For copper homeostasis and copper tolerance, copper-sensing two-component system response regulator CusR, copper resistance protein D, copper-translocating P-type ATPase (EC 3.6.3.4), copper homeostasis protein CusE/F were annotated. The study has also detected multiple heavy metal resistant genes (cobalt-zinc-cadmium resistance) viz. cobalt-zinc-cadmium resistance protein CzcA/D, DNA-binding heavy metal response regulator protein, cation efflux system protein CusA, and response regulator of zinc sigma-S4-dependent two-component system.

3.3. Stress response genes

The sites contaminated with industrial discharges are often found under stress of xenobiotic compounds. The functional capabilities of native microbial community for different stress response were characterized into nine distinct pathways and clusters as observed from Fig. 3. The figure showed the abundance of genes for oxidative stress pathway followed by osmotic stress and heat shock.

The bacterial communities habitually develop various mechanisms for countering the multiple stresses under such heterogeneous habitat. One of the most frequently encountered stresses is oxidative stress during normal cell metabolism as well as during degradation of xenobiotic compounds. The study has detected genes like catalase (EC 1.11.1.6), peroxidase (EC 1.11.1.7) and manganese superoxide dismutase (EC 1.15.1.11) involving in protection from reactive oxygen species and superoxide radicals, where former two genes were more abundant in the metagenome (Table S2/Stress Response). Other oxidative stress counteracting genes encoding for gamma-glutamyltranspeptidase (EC 2.3.2.2), glutathione synthetase (EC 6.3.2.3), glutathione S-transferase (EC 2.5.1.18), hydroxyacetylglutathione hydrolase (EC 3.1.2.6), glutathione reductase (EC 1.8.1.7) responsible for glutathione: biosynthesis and gamma-glutamyl cycle and glutathione: non-redox and redox reactions respectively were annotated. Glutathione S-transferase is also involved in metabolism of different drugs and degradation of xenobiotic compounds through cytochrome P450. Genes encoding for the enzymes required during countering the oxidative stress like choline dehydrogenase (EC 1.1.99.1), choline-sulfatase (EC 3.1.6.6), L-proline glycine betaine binding ABC transporter protein ProX/V (EC 3.A.1.12.1), aquaporin Z responsible for choline, betaine uptake and betaine biosynthesis and osmoregulation respectively were detected from the polluted sites.

Stress responsive genes are usually found in bacterial system. At polluted sites of Sachin Industrial Estate, stress related genes were mainly identified from Actinomycetales, Rhizobiales, Burkholderiales, Enterobacteriales, Rhodobacterales, Bacillales, etc. At the rank genus, mainly Mycobacterium sp., Streptomyces sp., Frankia sp., Bradyrhizobium sp., Mesorhizobium sp., Methylobacterium sp., Rhodobacter sp., Stenotrophomonas sp., Xanthomonas sp., Magnetotirpillum sp., Acidiphilum sp., Nitrobacter sp., Anabaena sp., Nostoc sp., Nitrosomonas sp., Anaeromyxobacter sp., Escherichia sp., Geobacter sp., Carboxydothromus sp., Burkholderia sp., Ralstonia sp., Bacillus sp., etc. were found to exhibit stress response genes.

3.4. Carbon metabolism

The functional capabilities for metabolism of carbon by native microbial community were characterized into twelve distinct pathways and clusters as observed from Fig. 4. The abundance of different genes for central carbohydrate metabolism, metabolism of one carbon compounds and di- and oligosaccharides were evident from the figure.

The soil ecosystem, whether is pristine or polluted, is rich in cellulose, lignin, chitin, etc. having abundant carbon sources derived from plant or other organisms (Liang et al., 2011). Genes for metabolizing these compounds were annotated from the metagenome. At the polluted sites, besides the profusion of various xenobiotic compounds, several oligosaccharides (and other disaccharides) were found. The metabolism of primary sources of electrons is highly essential for degradation of xenobiotic compounds. Maltodextrin glucosidase (EC 3.2.1.20) alternatively also known as alpha-glucosidase, which catalyzes the breakdown of starch and disaccharides to glucose monomer was annotated from the sequenced metagenome. Genes encoding for the enzymes like chitinase (EC 3.2.1.14), neopullulanase (EC 3.2.1.135), 3-alpha-amylase (EC

Fig. 3. Relative differential abundance of functional features for various pathways for stress related genes annotated from metagenome of polluted soils of Sachin Industrial Estate, Sachin. The functions were mapped on Subsystem database on MG-RAST server with a sequence similarity of ≥60% confidence level, having e-value cutoff of 1 x 10^-5. A: Miscellaneous; B: Acid Stress; C: Cold Shock; D: Desiccation Stress; E: Detoxification; F: Heat shock; G: Osmotic stress; H: Oxidative stress; I: Periplasmic Stress.

Fig. 4. Relative differential abundance of functional features for various pathways for metabolism of carbon annotated from metagenome of polluted soils of Sachin Industrial Estate, Sachin. The functions were mapped on Subsystem database on MG-RAST server with a sequence similarity of ≥60% confidence level, having e-value cutoff of 1 x 10^-5. A: Miscellaneous; B: Aminosugars; C: CO2 fixation; D: Central carbohydrate metabolism; E: Di- and oligosaccharides; F: Fermentation; G: Glycoside hydrolases; H: Monoosaccharides; I: One-carbon Metabolism; J: Organic acids; K: Polysaccharides; L: Sugar alcohols.
3.2.1.1) and glucoamylase (EC 3.2.1.3), periplasmic z-amylase for metabolism of chinin, starch and other polymers were detected. Different dehydrogenases (partly required for xenobiotic metabolism) with EC No. such as EC 1.1.1.94, EC 1.1.1.37, EC 1.1.1.42, EC 1.1.1.41, EC 1.2.4.1, EC 1.2.1.3 were annotated in significant number, suggested their active role in general cell metabolism and xenobiotic degradation (Table S2/Carbon Metabolism).

Other significant genes for carbon metabolism detected in high abundance were Methylmalonyl-CoA mutase (EC 5.4.99.2), Methylcrotonyl-CoA carboxylase carboxyl transferase (EC 6.4.1.4) Butyryl-CoA dehydrogenase (EC 1.3.99.2), decarboxylase, ethylmalonyl-CoA mutase, Glycogen debranching enzyme (EC 3.2.1.-), UDP-glucose 4-epimerase (EC 5.1.3.2) and many other genes. CoB-CoM heterodisulfide reductase (EC 1.8.98.1) and Formylmethanofuran dehydrogenase (EC 1.2.99.5) were detected, which are required during methanogenesis. Another genes detected for one carbon metabolism were 5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20), Formyltetrahydrofolate deformylase (EC 3.5.1.10) and Methylene-tetrahydrofolate dehydrogenase (NADP+) (EC 1.5.1.5).

The carbon fixation genes at the polluted sites were mainly identified from Actinomycetales, Bacillales, Oscillatoriales, Rhizobiales, Sphingomonaedales, Burkholderiales and others. The oligocharide, disaccharide and central metabolism of carbon compounds were universally present in all the microorganisms, still few of the dominantly observed group of bacteria exhibiting this ability were Actinomycetales, Bacillales, Burkholderiales, Enterobacteriales, Myxococcales, Rhizobiales, Rhodobacterales, Rhodospirillales which corresponds to the dominant community of the polluted sites of Sachin Industrial Estate. The one carbon metabolism ability was predominantly identified from Actinomycetales, Bacillales, Burkholderiales, Enterobacteriales, Rhizobiales, Rhodobacterales, Sphingobacteriales, Sphingomonadales and others. The genes responsible for above functions were mainly identified at genus level from Nocardia sp., Corynebacterium sp., Bradyrhizobium sp., Geoactinobacillus sp., Bacillus spp. Bordetella sp., Burkholderia sp., Cyanobacterium sp., Methylbacterium sp., Methylcella sp., Spirulina sp., Carboxythermus sp., Enterobacter sp., Escherichia sp., Photobacterium sp., Anaerobacter sp., Azothobacter sp., Acidithiobacillus sp., Chloroflexus sp., Methylbacterium sp., Rhizobium sp., Granulibacter sp., Rhodospirillum sp. and many more.

3.5. Nitrogen metabolism

Recycling and replenishment of elemental nitrogen and its free availability is highly essential for degradation of xenobiotic compounds and sustaining the biogeography of the polluted environment. The functional capabilities for metabolism of carbon by native microbial community were characterized into twelve distinct pathways and clusters as observed from Fig. 5. The abundance of genes for ammonia metabolism along with nitrate and nitrite ammonification is evident from the figure. It was observed that industrial effluents generally contain high concentration of nitrogenous compounds such as nitrate and nitrite and/or ammonia.

The study revealed that bacterial community at polluted site harbored several genes for ammonia assimilation, denitrification, dissimilatory nitrite reductase, nitrate and nitrite ammonification, nitrogen fixation and nitrosative stress. The annotated metagenome showed the presence of ammonium transporter protein, glutamate synthase [NADPH] (EC 1.4.1.13), glutamine synthetase GlnN (EC 6.3.1.2), assimilatory nitrate reductase (EC 1.7.99.4), nitrate/nitrite transporter, nitrite reductase [NAD(P)H] (EC 1.7.1.4), respiratory nitrate reductase (EC 1.7.99.4) among other various genes for nitrogen metabolism (Table S2/Nitrogen Metabolism).

The nitrogen metabolism genes were mainly identified from Actinomycetales, Alteromonadales, Burkholderiales, Enterobacteriales, Myxococcales, Planctomycetales, Rhizobiales, Rhodocyclales, Thermales and others. The bacterial genera playing major role in nitrogen metabolism at the polluted sites were Frankia, Mycobacterium, Alkalicoccus, Symbiobacterium, Burkholderia, Methylibium, Blastopirellula, Agrobacterium, Bradyrhizobium, Rhodopsedomonas, Thermus, Chitinophaga, Rhodospirillum, Azorarcus, Gloeobacter, Anaabana, Escherichia, Cyanoethec, Bacillus, Sulfurifydrogenibium and other genus.

3.6. Sulphur metabolism

Elemental sulphur in various forms (inorganic or organic) is commonly found in different concentrations at sites polluted by industrial discharges. Although sulphur metabolism is generally found under anoxic conditions, certain bacteria can assimilate sulphur/sulfate under oxygen rich environment having ability towards the protection from ROS and other free radicals (Abbai et al., 2012). The functional capabilities for metabolism of sulphur by native microbial community were characterized into ten distinct pathways and clusters as observed from Fig. 6. The figure revealed that abundance of genes for inorganic sulphur assimilation and alkanesulphonate assimilation along with sulphur oxidation as noted above several genes involved in protection from oxidative stress and reactive oxygen species were annotated from the sequenced metagenome from polluted sites of Sachin Industrial Estate. Arylsulfatase (EC 3.1.6.1), which generally removes sulfate (hydrolysis) from phenol rings and required for assimilation of alkanesulfonate were annotated along with the gene responsible for sulphur oxidation i.e. sulphur oxidase. For inorganic sulphur metabolism one of the significant genes sulfate adenyllytransferase (EC 2.7.7.4) was annotated along with sulfite reductase (EC 1.8.1.2). Another gene for assimilation of alkanesulphonate, Alkanesulfonate monoxygenase (EC 1.14.14.5) was also observed (Table S2/Sulphur Metabolism).

The sulphur assimilation and metabolism genes were mainly

The annotation of sequences for taxonomic assignment with MSNR, IMG and LCA were consistent throughout the metagenome and the databases. Therefore the reproducibility of the study can be authenticated and the results obtained are valid.

### 3.8. Estimating alpha diversity

For better understanding of the diversity, different diversity indices are commonly used, which reflects the species richness and species diversity of the given habitat. The estimation of diversity indices suggested that, the release of xenobiotic compounds at Sachin Industrial Estate, have not reduced the microbial species diversity as revealed by all three Simpson diversity indices from Table 1. Since (1-D: 0.9915) and (1/D: 117.67) and Shannon’s diversity index: 5.15 were comparatively higher, the richness in the diversity can easily be anticipated. To make these values more effective and logical, Shannon’s diversity index was converted into effective number of species (ENS) which provides the definite number of species observed at the polluted sites of Sachin Industrial Estate. The ENS for *H* of 5.15 was 172.43 and it can be observed from Table 1 that, the calculated ENS value was closer to the Fisher’s alpha diversity index of 155.7. The plot between number of expected OTUs and the number of reads i.e. Rarefaction curve (Figure S1) showed that, the curve is about to reach the asymptote state, hence majority of species diversity was covered which were well reflected by higher Chao1 index of 844.2.

### 3.9. Community synergism

The continuous evolutionary process and selective pressure have provided bacteria an ability to utilize wide range of compounds as electron acceptors depending upon the ecosystem they are exposed. They work as interdependent community and their functioning are highly intertwined with ecosystem dynamics. Based on the obtained results, we tried to recreate the community dynamics of abundant bacterial taxa considering their major role in metabolism of xenobiotic compounds in the polluted environment. We have revealed the heterotrophic and autotrophic strategy which supports the native community to colonize in polluted niches of Sachin Industrial Estate. The association of important biogeochemically metabolic genes (both for bioremediation and general metabolism) with their phylogenetic identity showed the adaptive behaviour of the microbial community towards environmental changes.

As observed above, important genes for carbon and nitrogen metabolism was associated with Actinomycetales, Bacillales, Burkholderiales, Oscillatoriales, Rhizobiales, Planctomycetales, Rhodocyclales, Myxococcales, Sphingomonadales. The niches specific (i.e. relevant to bioremediation) genes were also annotated from Actinomycetales, Bacillales, Burkholderiales, Rhizobiales, Rhodocyclales, Myxococcales, Sphingomonadales along with Pseudomonadales, Chloroflexales, Rhodobacterales. Moreover, the stress response genes were identified from the above mentioned taxa along with Rhodospirillales. Besides the functional phylogenetic characterization, the taxonomic identification was performed using 16S rRNA gene from the metagenome. The abundance of Actinomycetales, Burkholderiales, Enterobacteriales, Myxococcales, Planctomycetales, Rhizobiales (and many other taxa) correlates well with the functional identification. Therefore, it can be

---

**Fig. 6.** Relative differential abundance of functional features for various pathways for metabolism of sulphur annotated from metagenome of polluted soils of Sachin Industrial Estate, Sachin. The functions were mapped on Subsystem database on MG-RAST server with a sequence similarity of ≥60% confidence level, having e-value cutoff of $1 \times 10^{-6}$. A: Galactosylceramide and Sulfate metabolism; B: Sulfate reduction- associated complexes; C: Sulphur oxidation; D: Thiodoxygen-dissulfide reductase; E: Inorganic Sulphur Assimilation; F: Alkanesulfonate assimilation; G: Alkanesulfonates Utilization; H: L-Cystine Uptake and Metabolism; I: Taurine Utilization; J: Utilization of glutathione as a sulphur source.

**Fig. 9.** The microbial community was also dominated by *Planctomyces* spp., *Geobacter* spp., *Candidatus Koribacter* spp., *Sorangium* sp., *Myco bacterium* spp., *Anaeromyxobacter* sp., *Burkholderia* sp., *Bradyrhizobium* sp., and others. The rare bacterial populations were *Selenomonas* spp., *Natanaerobius* spp., *Desulfobacterium* spp., *Deferribacter* spp., *Thi monas* spp., *Candidatus puniceispirillum*, *Dehalogenimonas* spp., *Pheobacter* spp. (Fig. 10). Additional rare species were *Nitrococcus* spp., *Hydrogenobaculum* spp., *Oceanibulbus* spp., *Arcobacter* spp., *Pectobacterium* spp., *Hirschia* spp., *Coxiella* spp., and many more.

Based on the obtained results, we tried to recreate the community structure of polluted soil of Sachin Industrial Estate, which supports the native community to colonize in polluted niches of Sachin Industrial Estate. The ENS for *H* of 5.15 was 172.43 and it can be observed from Table 1 that, the calculated ENS value was closer to the Fisher’s alpha diversity index of 155.7. The plot between number of expected OTUs and the number of reads i.e. Rarefaction curve (Figure S1) showed that, the curve is about to reach the asymptote state, hence majority of species diversity was covered which were well reflected by higher Chao1 index of 844.2.

3.7. Bacterial taxonomy and phylogenetic affiliation

Profiling of taxonomic identification and phylogenetic affiliation of metagenomic sequences of microbial (especially bacterial) diversity and their relative abundance provided the insight into the community structure of polluted soil of Sachin Industrial Estate, Sachin. As mentioned above the microbial community was assessed by evaluating the similarities between the conserved families and domains of metagenomic sequences from (1) M5NR and (2) IMG and (3) LCA databases on MG-RAST server. The annotation revealed a unilateral dominance of eubacterial sequences in the community. At the level of order, where more diversity was observed, the abundance of Actinomycetales, Rhizobiales, Enterobacteriales and Entero bacteriales was observed along with the species of Enterobacteriales, Burkholderiales, Myxococcales and Planctomycetales (Fig. 8). The annotation also revealed that nearly 5% of the community at polluted sites of Sachin Industrial Estate, Sachin was unclassified.

The polluted sites of Sachin were predominantly occupied by *Streptomyces* spp., *Escherichia* spp., *Candidatus* spp., *Solibacter* spp., *Conexibacter* spp. (Fig. 9). The microbial community was also dominated by *Planctomyces* spp., *Geobacter* spp., *Candidatus Koribacter* spp., *Sorangium* sp., *Myco bacterium* spp., *Anaeromyxobacter* sp., *Burkholderia* sp., *Bradyrhizobium* sp., and others. The rare bacterial populations were *Selenomonas* spp., *Natanaerobius* spp., *Desulfobacterium* spp., *Deferribacter* spp., *Thi monas* spp., *Candidatus puniceispirillum*, *Dehalogenimonas* spp., *Pheobacter* spp. (Fig. 10). Additional rare species were *Nitrococcus* spp., *Hydrogenobaculum* spp., *Oceanibulbus* spp., *Arcobacter* spp., *Pectobacterium* spp., *Hirschia* spp., *Coxiella* spp., and many more.

### 3.6. Metabolic pathways

- **I:** Alkanesulfonates Utilization
- **J:** Taurine Utilization
- **K:** L-Cystine Uptake and Metabolism
- **L:** Inorganic Sulphur Assimilation
- **M:** Alkanesulfonates Assimilation
- **N:** Galactosylceramide and Sulfate metabolism

**Table 1.** Since (1-D: 0.9915) and (1/D: 117.67) and Shannon’s diversity index: 5.15 were comparatively higher, the richness in the diversity can easily be anticipated. To make these values more effective and logical, Shannon’s diversity index was converted into effective number of species (ENS) which provides the definite number of species observed at the polluted sites of Sachin Industrial Estate. The ENS for *H* of 5.15 was 172.43 and it can be observed from Table 1 that, the calculated ENS value was closer to the Fisher’s alpha diversity index of 155.7. The plot between number of expected OTUs and the number of reads i.e. Rarefaction curve (Figure S1) showed that, the curve is about to reach the asymptote state, hence majority of species diversity was covered which were well reflected by higher Chao1 index of 844.2.

### 3.3. ENS calculation

**Table 1.** Since (1-D: 0.9915) and (1/D: 117.67) and Shannon’s diversity index: 5.15 was 172.43 and it can be observed from Table 1 that, the calculated ENS value was closer to the Fisher’s alpha diversity index of 155.7. The plot between number of expected OTUs and the number of reads i.e. Rarefaction curve (Figure S1) showed that, the curve is about to reach the asymptote state, hence majority of species diversity was covered which were well reflected by higher Chao1 index of 844.2.

### 3.9. Community synergism

The continuous evolutionary process and selective pressure have provided bacteria an ability to utilize wide range of compounds as electron acceptors depending upon the ecosystem they are exposed. They work as interdependent community and their functioning are highly intertwined with ecosystem dynamics. Based on the obtained results, we tried to recreate the community dynamics of abundant bacterial taxa considering their major role in metabolism of xenobiotic compounds in the polluted environment. We have revealed the heterotrophic and autotrophic strategy which supports the native community to colonize in polluted niches of Sachin Industrial Estate. The association of important biogeochemically metabolic genes (both for bioremediation and general metabolism) with their phylogenetic identity showed the adaptive behaviour of the microbial community towards environmental changes.

As observed above, important genes for carbon and nitrogen metabolism was associated with Actinomycetales, Bacillales, Burkholderiales, Oscillatoriales, Rhizobiales, Planctomycetales, Rhodocyclales, Myxococcales, Sphingomonadales. The niches specific (i.e. relevant to bioremediation) genes were also annotated from Actinomycetales, Bacillales, Burkholderiales, Rhizobiales, Rhodocyclales, Myxococcales, Sphingomonadales along with Pseudomonadales, Chloroflexales, Rhodobacterales. Moreover, the stress response genes were identified from the above mentioned taxa along with Rhodospirillales. Besides the functional phylogenetic characterization, the taxonomic identification was performed using 16S rRNA gene from the metagenome. The abundance of Actinomycetales, Burkholderiales, Enterobacteriales, Myxococcales, Planctomycetales, Rhizobiales (and many other taxa) correlates well with the functional identification. Therefore, it can be
postulated that the core metabolism and biogeography of the polluted ecosystem are sustained by these community. It was observed that niche specific unique pathways are being shaped up by the acclimatized community in the micro-habitat depending upon the environmental pressure (Gianoulis et al., 2009).

4. Discussion

The growth of modern chemistry, innovations in structural dynamics has led to the development of more resistant compounds with indefinite structural diversity. The abundance of these
synthetic organic compounds in the open environment has been mobilized by the intense civilization (Ojo, 2007). Thus, the exposure of microbes to synthetic organic compounds from past two centuries have created selective pressure on bacteria/microorganisms for evolving the enzymatic cascade mechanism necessary for transforming or degrading such compounds (Nojiri et al., 2004). As noted above, Sachin Industrial Estate manufactures different assorted xenobiotic compounds of heterogeneous nature. The liquid and solid wastes, raw and unused products, intermediates, etc. are constantly being released into the nearby soil and water bodies over the years.

This contamination of an open environment is visible which immediately affects the native microorganism in the ecosystem. Microbes have evolved for more than 2.5 million years ago and are the most successful life forms on the earth, being able to colonize the most extreme and diverse biotopes (Lorenzo, 2001). They have

Fig. 9. Relative abundance of dominant phylotype (bacterial genera) annotated from metagenome of polluted soils of Sachin Industrial Estate, Sachin. The abundance is represented in terms of total effective bacterial sequences from the metagenome using MG-RAST server at a sequence similarity of ≥60% confidence level, having e-value cutoff of $1 \times 10^{-5}$. 
developed numerous mechanisms to draw energy from the diverse environment and can consume amazingly large pool of chemicals as carbon and energy sources (Lorenzo, 2001). Hence to understand the effect of these pollutants on life, it is necessary to analyze the metabolic capabilities evolved by the native microbial community for countering the adverse effect of xenobiotic compounds and to identify their taxonomic linkages for recognizing their role in such environment.

In this study, several different genes encoding for various enzymes required for degradation of aromatic compounds were
annotated. The catabolic potential of the community is to be identified for different genes and enzymes involved in the biodegradation of pollutants. Different enzymes, like mono-oxygenases, dioxygenases, dehydrogenase, reductases, ligases directly/indirectly involved in bioremediation were abundantly detected. The annotation of these genes indicated that the indigenous community has acquired the ability to catabolize various xenobiotic compounds. The study of enzymes in bioremediation has gained significant importance in the last decade which led to the development of different approaches for their further exploration. The enzymes actively involved in degradation are always found to have broad substrate specificity and have acquired different catabolic routes. Further, it was also observed that these enzymes are well linked to their protein phylogeny, but their taxonomic identification differs from the host organisms (Perez-Pantoja et al., 2009).

The obvious reason for their discretion is the enzymatic potential of bacterial community of contaminated environment of Sachin Industrial Estate was analyzed in terms of their adaptation for degradation of xenobiotic compounds along with the assimilation of carbon and nitrogen metabolism and resilience towards stress generated because of profusion of anthropogenic activities at the site. In this study genes encoding for different enzymes relevant to biodegradation/bioremediation were identified from the sequenced metagenome. Moreover, detection of genes encoding for enzymes from different arrays of reaction, operating in unregulated pathways for metabolism of synthetic compounds (along with the response towards stress and DNA repair system), corresponds to the fact that polluted sites of Sachin Industrial Estate receives industrial effluent from a spectrum of industries manufacturing diverse synthetic compounds.

The high abundance of these genes annotated from the sequenced metagenome indicated that indigenous community have well developed mechanism to counteract the harmful effect of xenobiotic compounds. Further, the abundance of such sequences may also suggest that industrial effluents frequently contains higher concentration of aromatic compounds, heavy metals, antibiotics/drugs, etc. The sequences were also in higher numbers for genes corresponding to oxidative and osmotic stress, which again suggest that effluent contains high concentration of unused salts and during metabolism of different xenobiotic compounds like aromatic acids and heavy metals, ROS are usually generated.

Until recently, studies were aimed to understand the enzymatic pathways or degradation potential of individual microorganisms. Like, studies on degradation of polyaromatic hydrocarbons by Kim et al. (2008), degradation of several aromatic compounds by Cupriavidus necator JMP134 (Perez-Pantoja et al., 2008) and degradation of polychloriantrene biphenyl (Pieper and Seeger, 2008, Denef et al. (2005, 2006) using three approaches demonstrated that, in Burkholderia xrenovorans LB400 there exist well coordinated three benzoate degradation pathways. However, these are the isolated cases, in nature microorganism always exists in form of community and their metabolism is always complexly coordinated, interconnected and highly regulated. So after the development of high throughput sequencing technologies, the focus has shifted towards studying the entire metagenome. Nonetheless, very few studies are being performed to understand the metabolic pathways and enzymes involved and the type of microbial communities responsible for biodegradation process at industrial sites. The study was mostly concentrated on sludge system of different CETPs, since the microbial communities are less complex than the polluted soil system. Therefore, the merit of this study is that an ability of the indigenous microbial community directly from soil ecosystem polluted by heterogeneous industrial discharges was studied.

In a different study from the metagenome of activated biomass from common effluent treatment plant (CETP), several genes were abundantly annotated for peripheral and central pathways for catabolism of aromatic compounds by More et al. (2014), in particular, the dominance of meta-cleavage pathway was observed. Different oxygenases required for degradation of aromatic compounds through central and peripheral pathways were annotated by Jadeja et al. (2014), with the abundance of 1,2-homogentisate dioxygenase and phenylacetate Co-A oxygenases. In a similar study from sludge samples of a petroleum refinery wastewater treatment system, large number of genes was annotated for phenol and biphenyl degradation (Silva et al., 2013). The study has identified the m-cleavage pathway for aerobic and anaerobic degradation of aromatic compounds (like phenol, benzoate, naphthalene, etc.).

The soil ecosystem, whether is pristine or polluted, is rich in cellulose, lignin, chitin, etc. having abundant carbon sources derived from plant or other organisms (Liang et al., 2011). Genes for metabolizing such compounds were annotated from the metagenome of Sachin Industrial Estate. Their metabolism always yield free electrons and in polluted environment they are required for catabolism of aromatic, heterocyclic, aliphatic and polyaromatic hydrocarbons, reduction of heavy metals and other xenobiotic compounds under aerobic conditions.

The microbial habitat of contaminated environment incessantly changes with the change in the effluent composition released at the sites with change in production cycles. Therefore, it is extremely important for microbes for quick and swift adaptation towards constantly changing environment and it requires timely and appropriate alterations in gene expression along with protein activity to counteract the altered conditions (Kazmierczak et al., 2005). The change in gene expression initiates at transcriptional level, in associations with different alternative sigma factors and core RNA polymerase for expression of set of new genes in response to stimuli signalling of changed environmental conditions (Kazmierczak et al., 2005). To support the notion we have detected the pre-dominance of RNA polymerase sigma factor along with the abundance of heat shock chaperone protein DnaK/J and intra-membrane protease RasP/YluC.
In any given ecosystem, certain group of organisms is successfully perpetuating and profusely found and can be termed as dominant community, having a major role in maintaining the biogeography of that biotope. While other organisms are relatively observed in less abundance, but they are also equally significant in the habitat and can be termed as rare community. These rare species have important role to play in the polluted habitat of Sachin Industrial Estate. This does not necessarily implies that the genus termed as ‘rare’ in this study can universally be accepted as rare, with the change in environmental conditions the species dynamics would also change. But, presently species such as Desulfbacterium spp. (chemolithotrophic, utilize inorganic sulphur, reduce sulphates to sulfides under anaerobic conditions) Delfrrbacter spp. (reduces sulphur, nitrate and arsenate), Nitrococcus spp. (nitrifiers, oxidizes nitrite to nitrate), etc. are important to dominant community for aiding in bioremediation/biodegradation. Few species metabolize carbon (viz. Selenomonas but under fermentative conditions, Natranaerobiobus are halophilic again under anaerobic conditions. Thus, results suggest that the metabolism of dominant community is creating the micro habitat which would be suitable for the perpetuation of rare phylotype and together both the communities are essential for sustaining the biogeography of polluted environment.

In our earlier study from Alang-Sosiya ship breaking yard, near Bhavnagar, due to chronic ship breaking activities, the marine environment was polluted with petroleum hydrocarbons (PHCs), polycyclic aromatic hydrocarbons (PAHs), heavy metals and other xenobiotics, where Proteobacteria and Firmicutes along with Bacteroidetes were predominantly found (Patel et al., 2014). Precisely, the community shift was evidently observed from Gammaproteobacteria at pristine environment to Betaproteobacteria and Epsilonproteobacteria at polluted sites. In a similar study, Desai et al. (2009) observed that because of prolong exposure of Cr(VI), bacterial community at such sites were shifted to Firmicutes from Proteobacteria at non-exposure site (pristine). Shah et al. (2013) observed that Pseudomonadales, Clostridiales, Bacteroidales, Actinomyceotae, Burkholderiales, Rhizobiales were dominant bacterial group at Kharicut canal, Vatva, Ahmedabad, having analogous scale of industrial pollution as of Sachin Industrial Estate. In this study also, the polluted sites of Sachin Industrial Estate, were dominated by Proteobacteria, Actinobacteria, Firmicutes. The analogous results form this study suggested that these groups of bacteria are well evolved for metabolism of xenobiotic compounds. Thus, indigenous microbial community of polluted siths would gradually adapt and evolved the mechanisms for successful perpetuation at industrially polluted environment.

Recently Muszynski et al. (2015) studied the microbial community dynamics of (a full scale) wastewater treatment plant and observed the dominance of β-proteobacteria, Actinobacteria and Chloroflexi across the three seasons. In a recent study, in wastewater treatment plant with anoxic-aerobic membrane bioreactor (MBR) with internal recirculation, Methylishaetales, Myxococcales, Rhodocyclales and phylum Plantomycetes were found to remove trace organic matters (besides total organic carbon and total nitrogen) (Phan et al., 2016). Khan et al. (2014) have enriched the microbial community at such sites were shifted to Firmicutes from Proteobacteria at non-exposure site (pristine). Shah and Madamwar (2013), obtained the enriched and acclimatized community of Pseudomonadales, Actinomycetales, Enterobacteriales, Bacillales, Rhizobiales and Actinomycetales. In the present study phylum beta-proteobacteria (Myxococcales, Rhodocyclales) were annotated for their role in metabolism of xenobiotic compounds. Pseudomonadales, Enterobacteriales, Bacillales, Rhizobiales and Actinomycetales have been detected for their role in metabolism of central and peripheral aromatic compounds and their intermediates.

5. Conclusion

Because of selective pressure of xenobiotic recalcitrant compound, bacterial competence for their degradation is constantly evolving which can be utilized for the removal of pollutants. As being hypothesized earlier, several genes encoding the enzymatic machinery required for degradation of xenobiotics were mapped (from the polluted metagenome of Sachin Industrial Estate) along with the accessory genes required to sustain the biogeography of polluted ecosystem. The autochthonous (bacterial) community at polluted sites was predominated by Actinobacteria and Proteobacteria. Therefore the study suggested that industrially polluted environment also contains rich bacterial community having genetic ability to degrade xenobiotic compounds.

Acknowledgements

Authors acknowledge Department of Biotechnology, Ministry of Science and Technology, New Delhi for financial support. ND gratefully acknowledges University Grants Commissions, New Delhi for providing Junior Research Fellowship.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibiod.2017.06.011.

References

nomics of aerobic bacterial degradation of aromatics. In: Timmis, K.N. (Ed.),
community dynamics in an anoxic-aerobic membrane bioreactor e Impact on
109, 61–72.
Poei ladres, G.J., Zandstra, M., Bosma, T., Kalakow, L.A., Larkin, M.J., Marchesi, J.R.,
isolated from geographically distinct locations possess a highly conserved gene
Pruesse, E., Quast, C., Knittel, K., Fucho, B.M., Ludwig, W., Pfeldes, J.O., Goclenk, F.O.,
2007. SILVA: a comprehensive online resource for quality checked and aligned
ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35 (21),
7188–7196.
diversity of plasmids bearing genes that encode toluene and xylene metabolism
in Pseudomonas strains isolated from different contaminated site in Belarus.
Shah, V., Madamwar, D., 2013. Community genomics: isolation, characterization and
Taxonomic profiling and metagenome analysis of a microbial community from
Silva, C.C., Hayden, H., Sawbridge, T., Mele, P., De Paula, S.O., Silva, L.C.F.,
Vidigal, F.M.P., Vicentini, R., Sousa, M.J.P., Torres, A.P.R., Santiago, V.M.J.,
Oliveira1, V.M., 2013. Identification of genes and pathways related to phenol
degradation in metagenomic libraries from petroleum refinery wastewater.
Springael, D., Ryngaert, A., Merlin, C., Toussaint, A., Meer, M., 2001. Occurrence of
Tn4371-related mobile elements and sequences in (chloro) biphenyl-
Stenuit, B., Eyers, L., Schuler, L., Agathos, S.N., George, I., 2008. Emerging high-
throughput approaches to analyze bioremediation of sites contaminated with
technol. 17, 200–204.
dichlorophenoxyacetic acid-degradative plasmids isolated from (chloro) biphenyl-
Vallaeys, T., Courde, L., McGowan, C., Wright, A.D., Fulthorpe, R.R., 1999. Phyloge-
nomic analyses indicate independent recruitment of diverse gene cassettes
during assemblage of the 2,4-D catabolic pathway. FEMS Microbiol. Ecol. 28,
373–382.
ecology and ‘omics’ technologies: towards understanding in situ biodegra-