## CONTENTS

**List of Tables**
i-ii  
**List of Figures**
iii-v  
**Abbreviations**
vi-ix  

### INTRODUCTION

1-4  

### 1. REVIEW OF LITERATURE  
5-44

1.1 Evolution and Development of Microbial Diversity  
5  
1.2 Contaminants of Metagenomic DNA and Extraction Methods  
7  
1.3 Microbial Diversity Analysis  
9  
1.3.1 Culture Dependent Methods  
9  
1.3.2 Culture Independent Methods  
10  
1.4 Sequencing  
12  
1.4.1 Amplicon Sequencing (Gene targeted Metagenomics)  
16  
1.4.2 Shotgun Sequencing  
18  
1.5 Use of Bioinformatics  
19  
1.5.1 Database Dependent Bioinformatics  
19  
1.5.2 Database Independent Bioinformatics  
22  
1.6 Functional Gene Centric Analysis  
23  
1.7 Microbial Diversity Case Study  
25  
1.7.1 Sediments  
26  
1.7.2 Xenobiotics  
27  
1.7.2.1 Trichloroethene  
28  
1.7.2.2 Hydrocarbons  
28  
1.8 Hexachlorocyclohexane(HCH)  
29  
1.8.1 HCH Contaminated Sites Across the Globe  
30  
1.8.2 HCH Dumpsites: The Indian Scenario  
34  
1.8.3 Biochemistry and Genetics of HCH Degradation  
37  
1.8.4 HCH Degrading- lin Genes  
38  
1.8.5 Toxic Effects of HCH Isomers  
40  
1.8.6 Microbial Community at HCH Dumpsite  
42
### 2. MATERIAL AND METHODS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Overview of Metagenomic Study</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>Site Selection</td>
<td>46</td>
</tr>
<tr>
<td>2.3</td>
<td>Physiochemical Test</td>
<td>47</td>
</tr>
<tr>
<td>2.4</td>
<td>HCH Residue Analysis</td>
<td>47</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Extraction of HCH Residues from Soil</td>
<td>48</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Residue Analysis on Gas Liquid Chromatography (GLC)</td>
<td>48</td>
</tr>
<tr>
<td>2.5</td>
<td>Soil DNA Isolation</td>
<td>49</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Soil DNA Isolation by Crude Method</td>
<td>49</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Soil DNA extraction by PowerMax Kit</td>
<td>50</td>
</tr>
<tr>
<td>2.6</td>
<td>Quantitation of DNA using Nanodrop</td>
<td>51</td>
</tr>
<tr>
<td>2.7</td>
<td>16S rRNA Gene Amplification from Soil DNA</td>
<td>52</td>
</tr>
<tr>
<td>2.8</td>
<td>Primer used in Diversity</td>
<td>52</td>
</tr>
<tr>
<td>2.9</td>
<td>Metagenomic Sequencing and Data Generation</td>
<td>53</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Tag Encoded FLX Amplicon Pyrosequencing</td>
<td>53</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Shotgun Sequencing</td>
<td>54</td>
</tr>
<tr>
<td>2.9.3</td>
<td>Illumina Sequencing</td>
<td>55</td>
</tr>
<tr>
<td>2.10</td>
<td>Metagenomic Diversity Analysis</td>
<td>55</td>
</tr>
<tr>
<td>2.10.1</td>
<td>Heat Map Construction</td>
<td>55</td>
</tr>
<tr>
<td>2.10.2</td>
<td>Metagenomic Small Subunit Ribosomal RNA (SSU rRNAs) Extraction</td>
<td>56</td>
</tr>
<tr>
<td>2.10.3</td>
<td>Qualitative and Quantitative Measurements of Phylogenetic Diversity</td>
<td>57</td>
</tr>
<tr>
<td>2.10.4</td>
<td>Tree Construction from 1000 Selected OTUs</td>
<td>58</td>
</tr>
<tr>
<td>2.10.5</td>
<td>Diversity Correlation by using Mantel Test</td>
<td>58</td>
</tr>
<tr>
<td>2.10.6</td>
<td>Rarefaction Curve</td>
<td>59</td>
</tr>
<tr>
<td>2.10.7</td>
<td>Diversity Indices</td>
<td>59</td>
</tr>
<tr>
<td>2.10.8</td>
<td>Environmental Gene Tags (EGT) Approach</td>
<td>59</td>
</tr>
<tr>
<td>2.11</td>
<td>Metagenomic Functional Analysis</td>
<td>60</td>
</tr>
<tr>
<td>2.11.1</td>
<td>CoMet</td>
<td>60</td>
</tr>
<tr>
<td>2.11.2</td>
<td>WebMGA ORF Prediction</td>
<td>61</td>
</tr>
<tr>
<td>2.11.3</td>
<td>Protein Function Annotation (PFAM)</td>
<td>61</td>
</tr>
<tr>
<td>2.11.4</td>
<td>KEGG (Kyoto Encyclopedia of Genes and Genomes)</td>
<td>62</td>
</tr>
<tr>
<td>2.11.5</td>
<td>COG (Cluster of Orthologous Grop)</td>
<td>63</td>
</tr>
<tr>
<td>2.11.6</td>
<td>Community Potential and Participation for HCH Degradation</td>
<td>64</td>
</tr>
</tbody>
</table>
3. RESULTS

3.1 Site for Soil Collection 67
3.2 Evaluation of HCH Residues from Soils 68
3.3 Physio-Chemical Properties of Dumpsite and Other Gradient Soils 68
3.4 Metagenomic DNA Isolation from Soil 69
3.5 Quantification of Soil DNA 70
3.6 16S rRNA Gene Amplification 70
3.7 Sequence Data Generation 71
3.8 Domain Percentage 71
3.9 Microbial Community Analysis by Tag Encoded FLX Amplicon Pyrosequencing (TEFAP) 72
3.10 Rarefaction Curves 73
3.11 Diversity Indices 74
3.12 Archeal Diversity 75
3.13 Fungal Diversity 75
3.14 HCH Degradating Anaerobic Bacteria 77
3.15 Genera Enriched at Dumpsite 77
3.16 Classification of Genera present in Dumpsite Soil Metagenome 78
3.17 Genera Enriched in Five km Datasets 79
3.18 Phylum Distribution Comparison by MG-RAST based upon GreenGenes (GG) and Ribosomal Database Project (RDP) Annotations 79
3.19 Diversity Correlation 80
3.20 Comparative Analysis of Microbial Diversity Obtained by Tag based and Metagenomic Approach (SSU rRNA) 81
3.21 The Environmental Gene Tags Method 82
3.22 Comparative Diversity Analysis Comparison Using All Three 16S Pyrotags, SSU rRNA and EGTs Approaches 83
3.23 Taxonomic Distribution by Comparing 454 Pyrosequencing and Illumina Sequencing 85
3.24 Comparison of 16S rRNA Variable Regions 86
3.25 Rarefaction Curve from Pfam Databases 87
3.26 Functional Correlation 87
3.27 Gene Centric Function Analysis 88
3.28 Cellular Processes Enriched over Dumpsite 89
3.29 Cellular Processes Enriched over all Gradients 90
3.30 $\text{lin}$ Genes Present in Metadata Sets 91
3.31 HCH Degradating Genes 93
3.32 $\text{lin}$ genes Diversity 93
3.33 Metabolic Pathways over Dumpsite 94

4. DISCUSSION 97-107

SUMMARY 108-109

REFERENCES 110-140

APPENDIX I

LIST OF PUBLICATIONS