Chapter 3

Materials and Methods for Wet Microbial Analyses

3.1 Sampling, DNA Extraction and PCR Amplification

Sludge samples were collected at different wastewater treatment plants treating both municipal and industrial wastewaters, and from laboratory scale reactors (SBR) operated under different experimental conditions. All the samples were stored at -80°C until use. In total, 64 activated sludge samples that were already sampled and stored, were used for the present study. Complete details (sample name, origin, type of sludge treatment) are given in Appendix I.
The samples are as follows,

- Municipal Wastewater treatment plant: 37 samples
  (36 from Kanto are and 1 from western part of Japan)
- Industrial wastewater treatment plant: 08 samples
- Laboratory Scale activated sludge reactor: 19 samples

Samples were dived into eight groups (A-H) and each sample was labeled from 1 to 8 and x with group name for the ease of handling. As pyrosequencing reactions are done on a plate on which 1 million beads are dispersed, by dividing the plate into plural sections (typically 2, 4, 8 or 16), one can analyze plural samples at one time. “A” to “H” corresponds to 8 regions on the plate for pyrosequencing. Further, as the author used 8 plus 1 barcodes (1 for no barcode sequence), for each of 8 regions, 9 samples can be analyzed. The sample names A1-A8, Ax, B1-B8, Bx, C1-C8, Cx, D1-D8, Dx, E1-E8, Ex, F1-F8, Fx, G1-G8. Gx, and Hx. Here, H1 to H8 was assigned samples not of mine but of another researcher’s.

Samples stored at -80°C were thawed and 50µl of the sample was diluted 20 times with Milli-Q water, sonicated by 250DA Advanced Digital Sonifier (Branson) with a special micro tip at an amplitude of 40% (20W) for 20 seconds.

The sonified liquid was further diluted for adjustment of template DNA concentration to 1- 10pg/µl for PCR. After dilution, PCR reaction was performed with a Thermal Cycler Dice (Takara, Japan).

PCR was carried using universal primers pairs, (Lane, 1991)

- 27forward / 519reverse

Primers are barcoded with 4 base fragment at -5’ end.

Eight different barcodes plus one without barcode were used for amplification of nine samples in one group.

Table 3.1 Sequences of Barcoded Primer Pair Sequences used.

Number	Barcode Sequence	Forward Primers (27f)	Reverse Primer (519r)
1	AAAA	***AAAA***-AGAGTTTGATCMTGGCTCAG	***AAAA***-GWATTACCGCGGCKGCTG
2	AATT	***AATT***-AGAGTTTGATCMTGGCTCAG	***AATT***-GWATTACCGCGGCKGCTG
3	ATAT	***ATAT***-AGAGTTTGATCMTGGCTCAG	***ATAT***-GWATTACCGCGGCKGCTG
4	ATTA	***ATTA***-AGAGTTTGATCMTGGCTCAG	***ATTA***-GWATTACCGCGGCKGCTG
5	TTTT	***TTTT***-AGAGTTTGATCMTGGCTCAG	***TTTT***-GWATTACCGCGGCKGCTG
6	TTAA	***TTAA***-AGAGTTTGATCMTGGCTCAG	***TTAA***-GWATTACCGCGGCKGCTG
7	TATA	***TATA***-AGAGTTTGATCMTGGCTCAG	***TATA***-GWATTACCGCGGCKGCTG
8	TAAT	***TAAT***-AGAGTTTGATCMTGGCTCAG	***TAAT***-GWATTACCGCGGCKGCTG

PCR reaction Mixture (per sample),

10X buffer	2.5µl
dNTP	2.0µl
ExTaq Polymerase	0.125µl
Primer (Forward)*	0.5µl
Primer (Reverse)*	0.5µl
Sample	2.5µl
ddH2O	16.875µl
Total Volume	**25.0 µl**

* Primers are barcoded with 4base fragment at -5’ end.

Barcoded primer sequences are given in Table 4.1

Each sample amplified with 8 replicates

Thermal program was set as follows,

- Initial heating to activate enzyme: 95°C for 600 seconds,
- 30 cycles of denaturation, annealing and extension:, 55.3°C for 30s and 72°C for 30s, respectively
- Final extension: 72°C for 600s

PCR product concentration was determined by Picogreen (Quant-iTTM PicoGreen dsDNA Reagent and Kits) from Invitrogen according to manufacture's instructions. Then, the PCR products were purified by QIAQuick PCR purification KIT according to manufacturer's instruction. Purified product concentrations were measured by UV absorbance using Nanodrop 1000 (manufactuer).

And the quality of the products was confirmed by running on a 2% Agrose gel (Electrophoresis). Purified PCR products from the samples, the same DNA amount, were mixed together to satisfy the minimum requirements for pyrosequencing (1µg of total DNA). Then, samples were submitted for pyrosequencing, pyrosequencer 454 Titanium (Roche, 454 Life Sciences, Connecticut, USA).

The pyrosequencing was done by Center for Omics and Bioinformatics/Department of Computational Biology, Graduate School of Frontier Sciences, The University of

Chapter 4

Development of 16S rDNA Pyrosequencing Workflow

4.1 Introduction

In the present chapter, the author attempted to develop a convenient workflow for 16SrDNA pyrosequencing data analysis.

The data used for development of workflow was obtained by the author, as described in section 4.2. In short, DNA was extracted from 64 samples, and partial 16S rRNA gene was amplified with barcoded primers. Then, PCR products were mixed, and were analyzed by a Roche 454 pyrosequencer. As a result, about 700,000 reads were obtained.

In 4.3, the author selected reads from 2 of the samples described in section 4.2. These samples were from activated sludge samples that were incubated in parallel with and without the addition of "activated sludge extract". The author worked on these samples because the number of reads to analyze was less. The author worked on the reads in a more manual way aided by "RDP pyrosequencing pipeline" a pyrosequencing data processing tool on the internet. Here is the summary for the pyrosequencing data used.

In 4.4, the author selected approx. 13,000 reads from 8 of the samples described in 4.2, and compared the microbial population structures in these samples. The author introduced QIIME to analyze pyrosequencing data from these samples. But the authors also used RDP Classifier to visualize phylum/class level compositions, used ARB to draw phylogenetic tree, and used FileMaker to draw heatmap at OTU level. As a whole, while sequence analysis of plural samples was made easy by QIIME, the author had to use different tools to prepare figures to compare microbial populations.

In 4.5, 100, 000 reads from 20 samples were used. These samples were from full scale wastewater treatment plants, three of them from small scale and the rest from large scale wastewater treatment plants. Here, the authors tried to compare microbial population structures in small and large scale treatment plants. The author introduced OTUMAMi, a software developed on FileMaker Pro, to help handle and summarize the data from QIIME.

Table 4.1 Comparison on three case studies performed.

Main Characteristics	Study 1	Study 2	Study 3
Number of Samples	2	8	20
Number of Sequences	4,495	~ 13,000	~ 100,000
Objectives	Compare microbial community compositions Compare T-RFLP data and Pyrosequence data	Compare microbial community compositions Compere DNA extraction methods	Compare microbial community compositions
Sequence Quality Analysis	RDP	RDP	QIIME
Taxonomic Classification	RDP	RDP	QIIME
Quantification of communities	Manual	Manual & Computational	Computational
Data Interpretation	MS Excel Charts	MS Excel Charts Heatmap (Computational module)	MS Excel Chart Heatmap (Computational Module :OTUMAMi) for Family and OTU level
Phylogenetic Analysis	ARB	ARB	Full & Partial Trees By QIIME

Case study 1 and 2

Conducted to develop a pathway for microbial community identification including data analysis tools and interpretation methods.

Case study 3

Performed to explain and confirm the functionality of developed computational module. The basic requirements for a microbial community analysis were clearly discussed in this case study.

4.2 Pyrosequencing Data

Outline of the pyrosequence data obtained,

In total, 734976 sequences were obtained by pyrosequencing. In order to grasp the outcome, here, the resulted sequences were checked for their quality by

1. identification of barcode attached fragments, and
2. extraction of fragment length greater than 300bp in length.

The sequences were sorted out based on the sequences of the barcode by using "RDP Pipeline Initial Process". **C**riteria for sorting 27f- 519r reads was:

1. Forward and Reverse primer maximum edit distance = 2
2. Sequence length greater than or equal 300bp

And the outcomes provided were,

- The number of reads matched to respective barcodes (from 1 to 8)

- Average length distribution by histogram
- Sequences left without matching

As shown in Table 4.2 and Fig. 4.2, the total number of sequences allocated for each barcode and sequences remained without tagging separately given. The trimmed sequence data were returned as "FASTA" formatted text files. In Table 4.2

and Fig. 4.2, "No tag" is meant for reads which didn't have barcode sequences. Such sequences are either from Ax, Bx, …, Hx samples for which primers without barcode were used, or sequences in which barcode region was anyhow partially damaged.

Fig.4.2 is the graphical interpretation of counts of trimmed sequence as per each barcode on sample basis. This helps to identify the number of barcodes successful in providing high quality fragments with appropriate length. Therefore, barcode 1, 2, 3 (Except Sample B3) and 4 were successful in generating quality fragments while Barcode no. 5 was not successful and 6, 7, 8 were poor.

In Fig. 4.2, size distribution of reads are shown. The size of the reads were mostly from 370bp to 500bp.

Complete read counts and size distribution patterns are reported in Appendix II for all sample groups.

List of Barcodes,

1- AAAA
2- AATT
3- ATAT
4- ATTA

5- TTTT
6- TTAA
7- TATA
8- TAAT

Table 4.2 Sample Group A: Number of sequences reported and size of reads based on barcode.

Sample	Bar Code	Total	Selected	Avg. Size (bp)
A	AAAA	17728	959	429
	AATT	24109	1142	429
	ATAT	20206	1011	435
	ATTA	11971	550	440
	TTTT	2823	166	409
	TTAA	4424	281	378
	TATA	3670	180	437
	TAAT	3321	131	426
No tag		33535	1288	447
	Total	121787	5708	

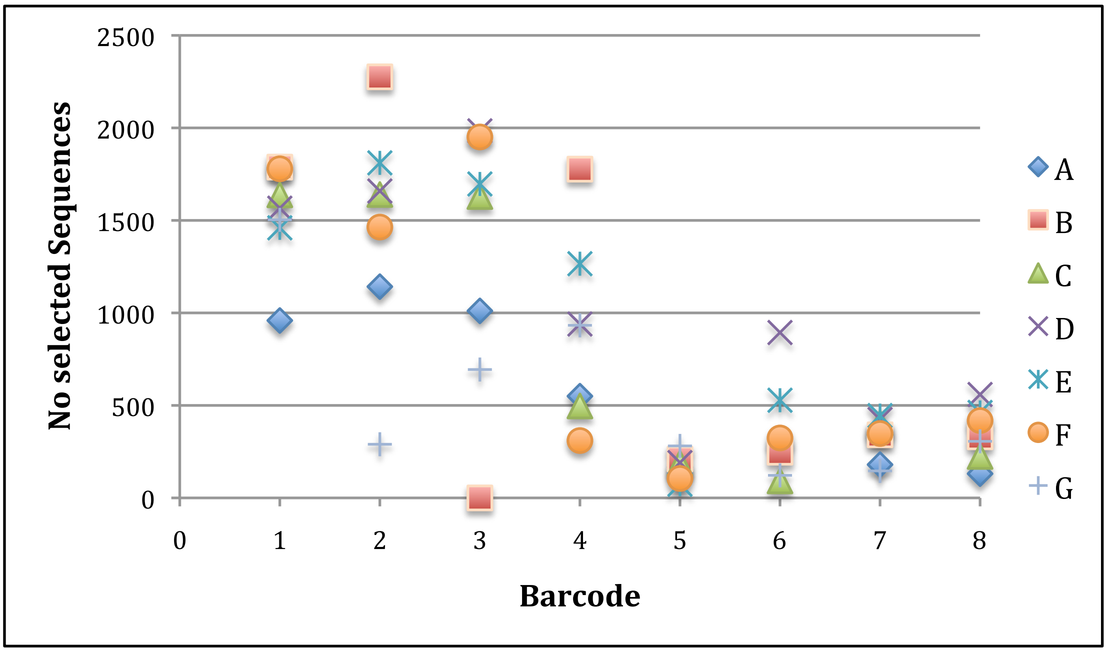
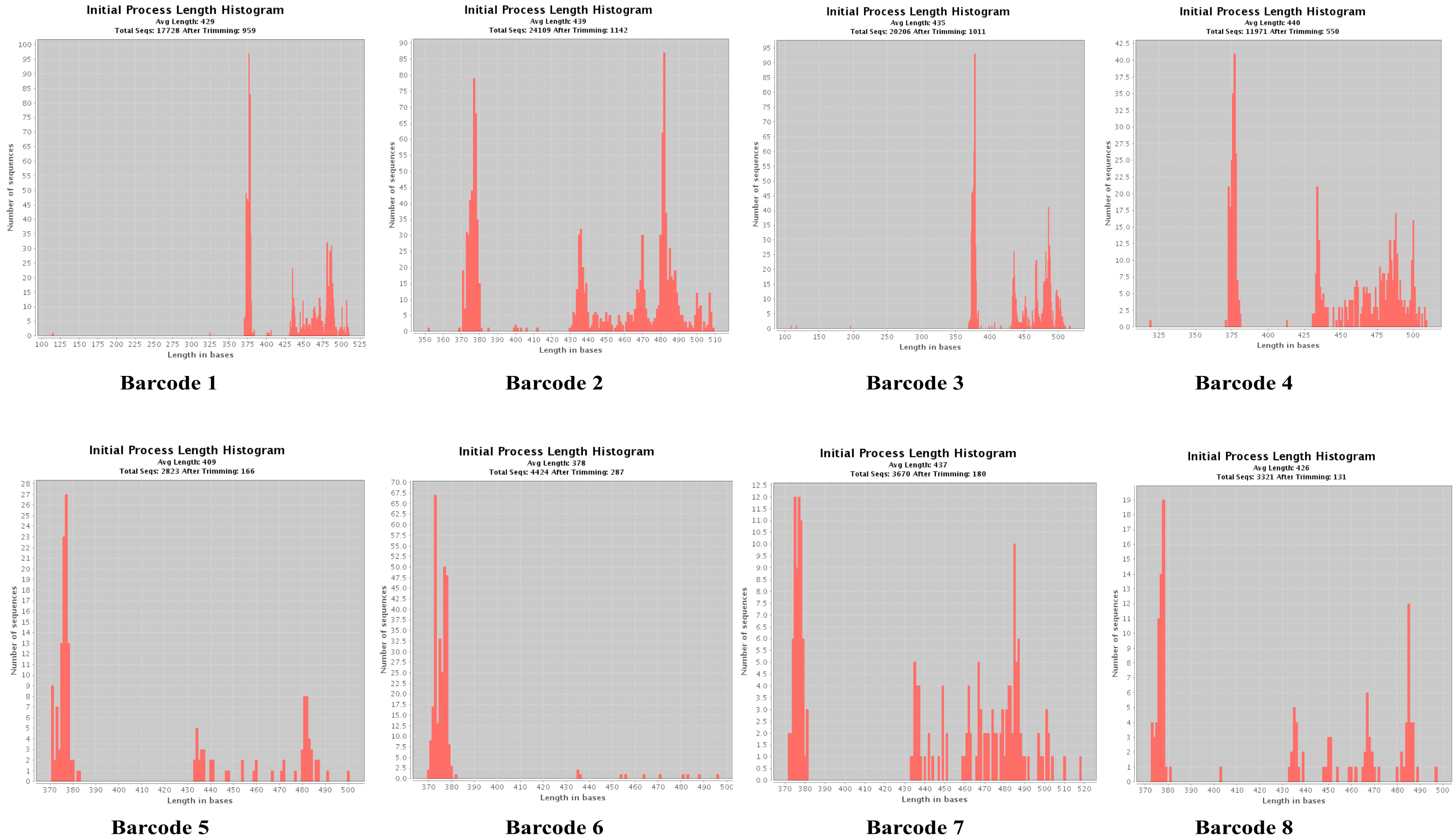


Fig. 4.1 Number of trimmed sequences obtained on each barcode matching for each sample.



Barcode 1 **Barcode 2** **Barcode 3** **Barcode 4**

Barcode 5 **Barcode 6** **Barcode 7** **Barcode 8**

Fig. 4.2 Fragment size (bp) distribution in eight barcodes for Sample A

4.3 Study 1: Pyrosequencing aided identification of microorganisms that were affected by activated sludge extract

<u>Background</u>

Two of the activated sludge samples reported in Satoh et al. (2009) were used in the present study for the identification of microbial community structure affected by the addition of activated sludge extract. They divided their activated sludge into two groups, and they incubated them under the same condition except that to one of them extract of activated sludge was added, and to another not. After 5 days of incubation, they collected the sample from both activated sludge, and analyzed microbial population structure in them by PCR/T-RFLP with 27f/519r primer set (fluorescent marker on the 5'- end) and *Hha*I restriction enzyme. They found clear differences of microbial population in these two activated sludge samples, as is shown in Fig. 4.1, where sample X is meant for the activated sludge incubated with activated sludge extract, and sample Y without. They concluded that there are chemicals in activated sludge extract that affect microbial population structure.

In the present study, the author put the focus to clarify which species in their activate sludge was affected by the addition of activated sludge extract.

In addition, the pyrosequencing outcomes should be comparable to those with T-RFLP, if the same region of the same gene is the target. Once pyrosequencing data is obtained, the expected size of restriction fragments can be computationally calculated, and thus, the real (experimentally obtained) T-RFLP pattern and in silico calculated T-RFLP pattern from pyrosequencing data can be compared. Both Satoh et al. (2009) and the present study used 27f/519r primer set, the outcomes from these studies can be compared. So, the second objective was to compare observed T-RFLP pattern and calculated T-RFLP patter based on pyrosequencing data.

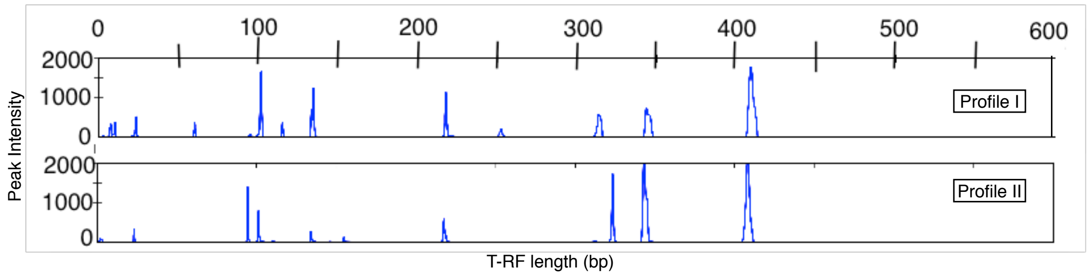


Fig.4.3 T-RFLP profiles obtained in the first set of experiments for sample X and Y (from Satoh et al., 2009)

Materials and Methods

In total, 4495 sequences were obtained for samples X and Y.

Sample X : 1279 (27f) + 1079 (519r) = 2358

Sample Y : 1165 (27f) + 972 (519r) = 2137

As shown in Fig. 4.2 the methodology comprised of three main sections,

1. Taxonomic assignment

   The sequences were classified with RDP classifier tool in RDP 10 database to assign the associated taxonomic groups.

2. Comparison of observed and calculated T-RFs

   The sizes of terminal-labelled restriction fragments (T-RFs) for the reads obtained by pyrosequencing were calculated by using FileMakerPro V10. The lengths of the calculated fragment sizes from 5' end and their abundances were compared with the actual T-RFLP profiles shown in Fig. 4.3.

3. Combination of T-RFs and identified microbial communities and interpretation by phylogenetic tree

   The sequences were imported to ARB software environment (Version. 5.1), aligned by ClustalW fast DNA alignment algorithm, and then a phylogenetic tree was build by the neighbour joining method. The tree was further edited to illustrate the main taxonomic groups related to each T-RFs.

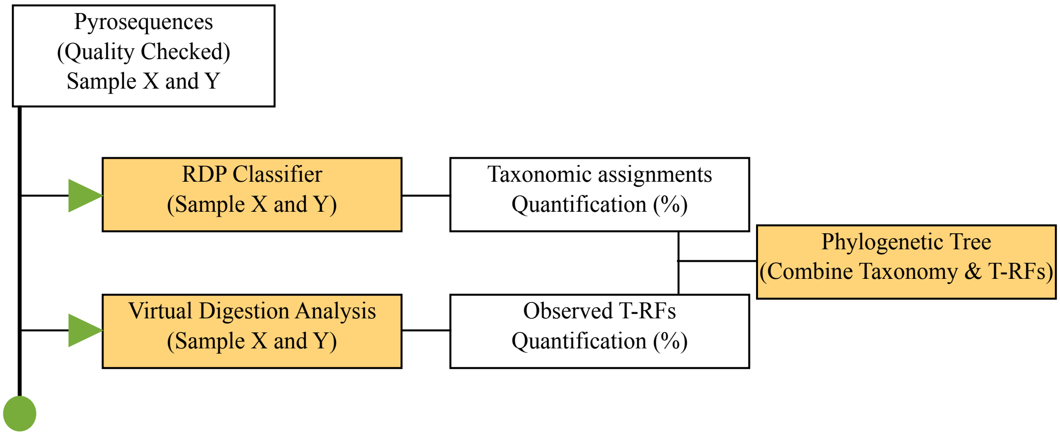


Fig.4.4 Workflow for case study 01

Results and Discussion

Selection of major microbial communities at class level was done by FileMaker Pro V.10 (FileMaker, USA) database management tool and extracted the taxonomic identities together with quantities.

Class level compositions were calculated for Samples X and Y, and are presented in Fig. 4.3. To prepare Fig. 4.5, output from RDP classifier was manually handled using Excel(Microsoft, USA) and text editor (mi) (MimikakiProject) software.

For each read, its T-RFs size was calculated using "Position" function in FileMaker Pro to know the position of the restriction site for *Hha*I restriction enzyme (GCG|C). Most of the reads were grouped into 10 groups, as shown in Table 4.3. The number of reads from samples X and Y for each of the 10 groups were counted manually.

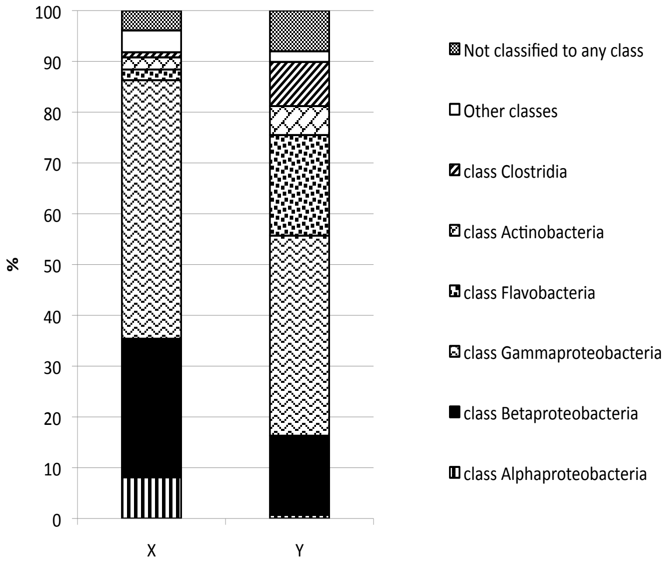


Fig.4.5 Class level classification for reads

Table 4.3 Grouping of reads based on expected and observed T-RFs sizes

Group	Calculated T-RF size (bp)	Observed T-RF size (bp)	No. of Reads X	No. of Reads Y	Enhancement X/Y
A	60-61	55 -60	14	36	0.35
B	85-86	85-90	15	30	0.45
C	143-145	135 - 140	29	104	0.25
D	148-150	-	49	103	0.43
E	205-206	200 -208	158	179	0.80
F	207-210	-	45	68	0.60
G	211-212	-	20	23	0.79
H	213-216	211-212	558	236	2.14
I	225-226	240 - 245	15	-	-
J	339-342	325 - 340	24	128	0.17

The phylogenetic relationships of the reads were analyzed by using ARB, and the result was presented as a phylogenetic tree. But with the original names (numbering) of the reads, it was difficult to see how each of the groups presented in Table 4.3 are distributed in the tree. So, the author found it is necessary to give proper names to the reads. Since, the number of reads were quite large, manual labeling of each sequence with it's T-RFs took lots of time. However, labeling to sequences were done efficiently by FileMaker programme. By using the names, the phylogenetic tree was drawn as shown in Fig.4.5.

In phylogenetic tree,

- Microbial communities Positively effected: Highlighted in "Green" color

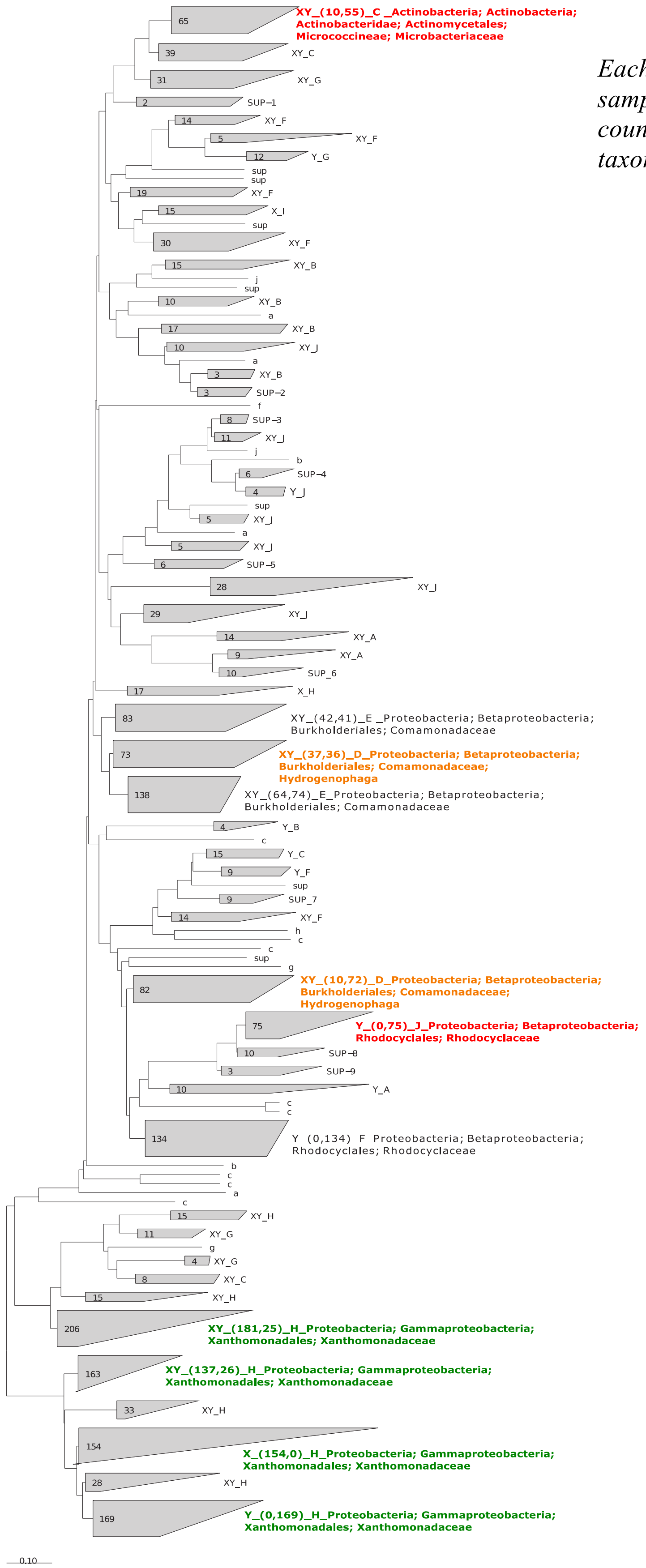
Microbial community population that enhanced by addition of activated sludge extract.

- Microbial communities Negatively effected: Highlighted in "Red" color

Microbial community population that reduced by addition of activated sludge extract.

- Microbial communities with Neutral effect: Highlighted in "Orange" color

Microbial community population didn't change by addition of activated sludge extract.


65
XY_(10,55)_C _Actinobacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Microbacteriaceae
39
XY_C
31
XY_G
2
SUP–1
14
XY_F
5
XY_F
12
Y_G
sup
sup
19
XY_F
15
X_I
sup
30
XY_F
15
XY_B
j
sup
10
XY_B
a
17
XY_B
10
XY_J
a
3
XY_B
3
SUP–2
f
8
SUP–3
11
XY_J
j
b
6
SUP–4
4
Y_J
sup
5
XY_J
a
5
XY_J
6
SUP–5
28
XY_J
29
XY_J
14
XY_A
9
XY_A
10
SUP_6
17
X_H
83
XY_(42,41)_E _Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae
73
XY_(37,36)_D_Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Hydrogenophaga
138
XY_(64,74)_E_Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae
4
Y_B
c
15
Y_C
9
Y_F
sup
9
SUP_7
14
XY_F
h
c
c
sup
g
82
XY_(10,72)_D_Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Hydrogenophaga
75
Y_(0,75)_J_Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae
10
SUP–8
3
SUP–9
10
Y_A
c
c
134
Y_(0,134)_F_Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae
b
c
c
a
c
15
XY_H
11
XY_G
g
4
XY_G
8
XY_C
15
XY_H
206
XY_(181,25)_H_Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae
163
XY_(137,26)_H_Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae
33
XY_H
154
X_(154,0)_H_Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae
28
XY_H
169
Y_(0,169)_H_Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae
0.10

Each microbial group labeled by sample of origin, corresponding read counts, calculated T-RFs group and taxonomic identification

Fig.4.6 Phylogenetic tree representation for pyrosequence reads

Limitations and Recommendations

1. The use of external platform for data analysis ideal for small sequence sets only targeting microbial community identification and simple quantification. These online tools can mange limited in number of sequences and sequences data import and export has to be done by the user. Therefore, it may cause some errors in sample management. However, in the present study only two samples with nearly 4,500 reads were used. Therefore, results can easily calculate manually and interpret directly as seen in Fig. 4.4 and Table 4.3.

2. Another limitation of RDP based classification is generation of text based data that need some editing before use. In order to calculate the community compositions based on obtained results some unwanted data to delete by user manually.

3. The presence of FileMaker tool manual handling was reduced and helped calculate microbial community composition for sample X and Y. for all samples in single layout. Therefore, having a computational module will be beneficial for the management of large data sets.

4. Comparison of expected T-RFs showed a clear correspondence with the actual T-RFs from profile reported in the previous study. Based on calculated T-RFs 10 groups were identified but T-RFLP profile did not show any peak pattern for some of the T-Rfs found from pyrosequence reads (ex. 148- 150 and 211-212). Even the obtained and calculated T-Rfs showed a good correspondence more detailed insight achieved by pyrosequencing reads.

Finally, a simple methodology was used in present study introduce novel data interpretation (Fig.4.5 T-RFs based Phylogenetic tree) that partly managed by semi-computational approach. A few computational tools like, FileMaker Pro applications, ARB and text editor tools were also introduced for better data management.

4.4 Study 2: Comparative study of bacterial communities in wastewater treatment plants by pyrosequencing of partial 16S rRNA gene

Background

Eight different activated sludge samples reported in section 4.2 were used for the present study. Five of the samples were from urban wastewater treatment plants (WWTP1-WWTP5) and another from a night soil treatment plant (WWTP6). As a comparison, in two of the samples DNA were extracted by Fast DNA Spin Kit for Soil (samples marked with b).

In the present study, Pyrosequencing was applied to clarify and compare bacterial population in activated sludge from different wastewater treatment plants.

Materials and Methods

Around 13,000 sequences in total were obtained from eight samples,

WWTP1 : 1162	WWTP5 : 1508
WWTP2 : 2953	WWTP5_b : 635
WWTP3 : 2203	WWTP6 : 1025
WWTP4 : 2014	WWTP6_b : 1168

As described in Fig. 4.6 the methodology comprised of two sections,

1. Taxonomic Assignment

   RDP classifier tool in Ribosomal Database Project 10 (RDP) used for assigning associate taxonomic groups.

2. Calculation of microbial community composition

   The percentage for sequences classified into each taxonomic class and their averages were calculated for all samples using FileMaker application. The calculated data import and graphically interpreted by MS Excel charts.

The phylogenetic tree (NJ method) was drawn by importing sequences to ARB software. The major microbial communities were highlighted and their associated read counts were included by editing tree by Inkspace (//) software.

3. QIIME based data interpretation

   The sequences were analysed by QIIME pyrosequencing platform and obtained the microbial compositions at Class level. The results were import to FileMaker application and heatmap color representation was developed to present the microbial composition change on sample basis.

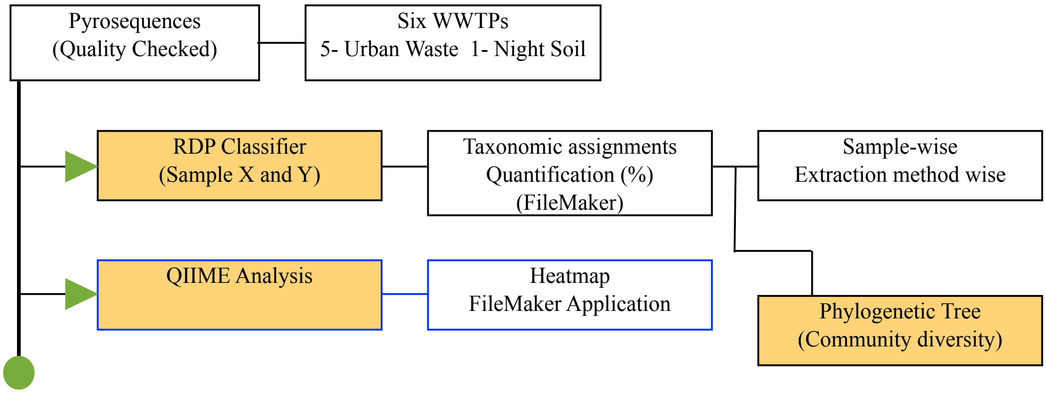


Fig.4.7 Workflow for case study 2

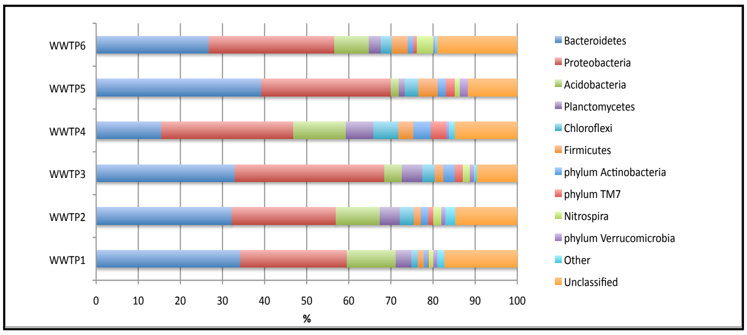
Results and Discussion

The results obtained from taxonomic aasignment and calculation of microbial community compositions, sequence reads were assigned to ten Phyla and 13 Classes mainly.

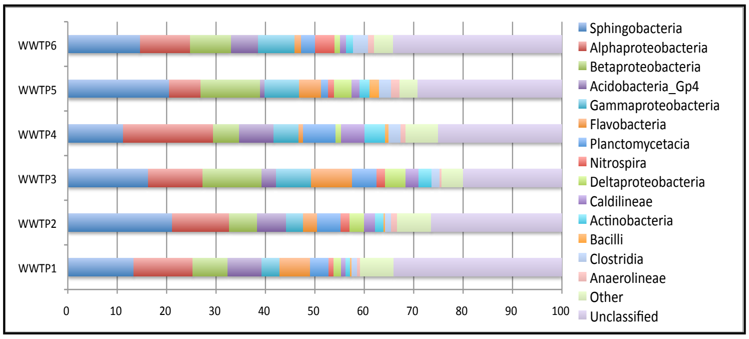
As seen in Fig. 4.8 and 4.9 the compostion of microbia communities at their Phyla and Class levels were clearly shown for each sample. And also, major and minor groups were clearly identified. The reads without assigned to any taxonomic affiliation were grouped as unclassified bacterial communities their percentages were also calculated and presented in Fig. 4.8 and 4.9.

For comparison, microbial population change in urban watewater and night soil samples and DNA extraction methods were easily compared by this way of taxonmic classifiication and graphically shown in chart format as in Fig. 4.9 and 4.10.

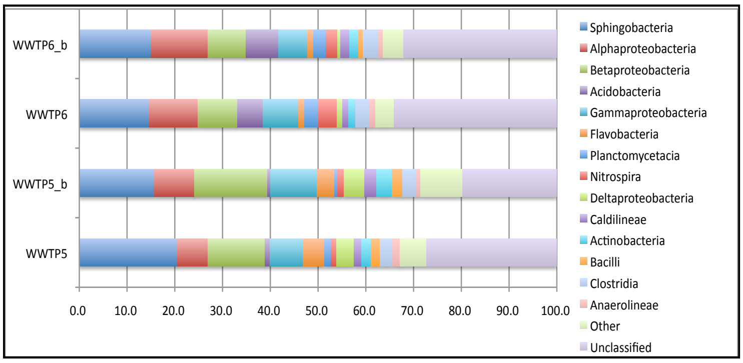
The diversity of microbial communities identified at their Class level was presented by phylogenetic tree (Fig. 4.11). The major microbial communities were highlighted with colors and labelled with the number of sequnce reads assigned to each group.



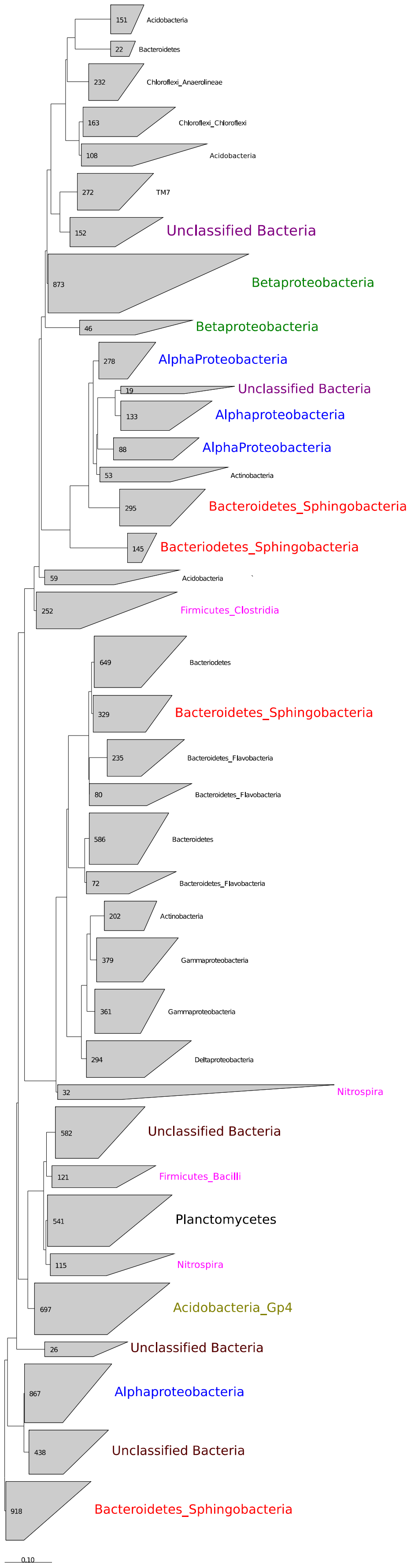
4.8 Taxonomic Assignments of reads at Phylum level



4.9 Taxonomic Assignments of reads at Class level



4.10 Taxonomic Assignments of reads form different DNA extraction methods (Class level)



* Major microbial communities highlighted with different colors and the number of reads assigned for each community labeled on the branch of the tree.

Fig. 4.11 Microbial community diversity classified and represented at Class level.

For the first time, QIIME introduced as a data analysis for pyrosequencing data. The sequence compositions obtained at their Class level assignment was used to produce the heatmap representation. The data import to FileMaker application and based on sequence percentages colors were assigned.

The development of heatmap (Fig. 4.12) by computational method was achieved and successfully interprets the composition changes on sample basis.

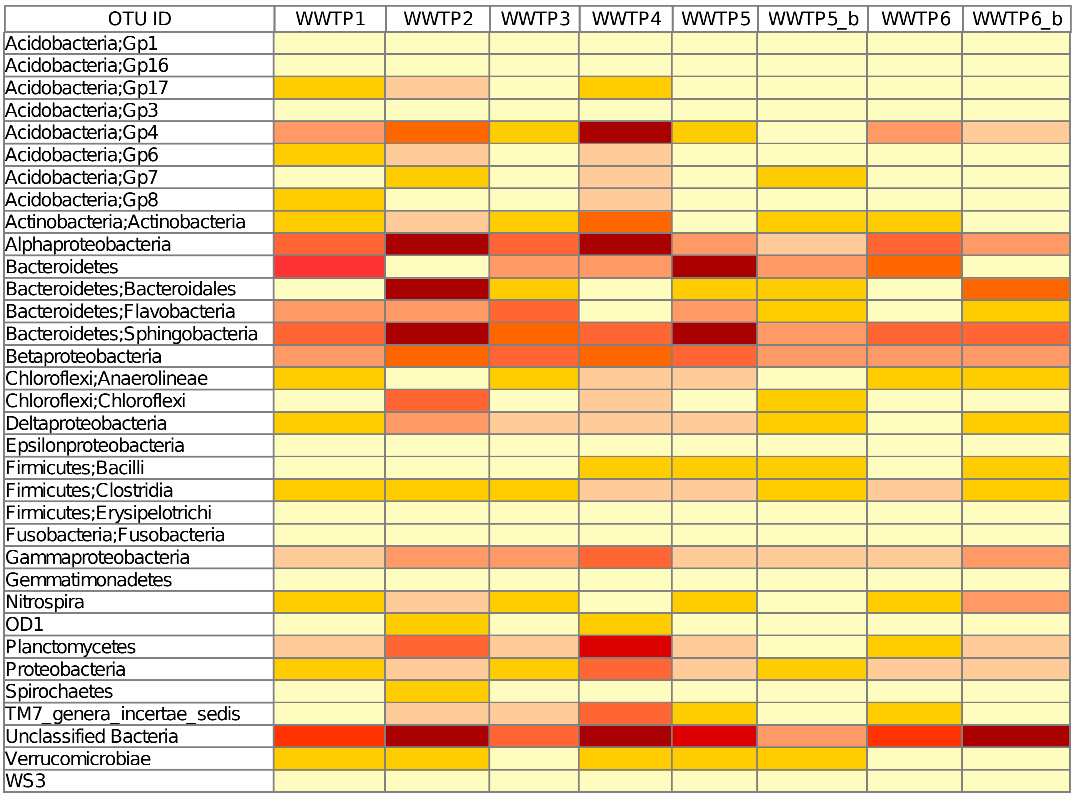


Fig.4.12 Heatmap representation on Class level taxonomic assignment for microbial communities found in six different wastewater treatment plants.

Limitations and Recommendation

1. The RDP based microbial community classification was applied in the present study. There user needs to upload sequences on sample basis and import results separately. Imported result given in text based which to be edit and extract the relevant statistical data. Therefore, for application of large set of sequences and samples, it will take a lot of time for manual import and export. And, also it can cause errors in sample labeling.

2. Compared to study 1, data analysis became quite easier in the present study. The use of FileMaker application helps for computing percentage of read assignments for several samples and storage of large number of reads. The time spend on calculation became less and easy presented in FileMaker layout. Therefore, manual calculation using spreadsheet application is omitted.

3. In the previous and present study, the microbial community compositions were mainly explained at Phylum and Class using MS Excel charts. Family, Order or Genus level assignment can be obtained by RDP classifier, but the text based result has to edit finely to import to FileMaker application. However, the QIIME data analysis resulted in OTU level community identification and it can useful to describe the community compositions at higher phylogenetic levels.

4. QIIME process rapidly analysed the sequence reads used in the present study and generates text based data file that can import to FileMaker application with less modification. Therefore, authors recommended QIIME for future studies.

Therefore, in study 3, QIIME data analysis process was completely applied for samples collected from 12 different wastewater treatment plants with different scale in operation. And, data representation mainly handled by a computational tool (OTUMAMi) developed by FileMaker software.

4.5 Study 3: Computational approach for revealing microbial community structures in large and small scale activated sludge systems

Background

In the present study, 20 activated sludge samples collected from 12 different wastewater treatment plants were used. All the samples were reported in section 4.2. Three samples (SA, SB and SC) from oxidation ditch processes less than 25,000 population and 1000ha of serviced area were grouped as small scale plants, while others as large scale plants (LA through LI). And sludge samples collected from different operational condition at the same treatment facility were also indexed (LA1-LA3) for ease of reference. Details of each wastewater treatment plants employed for the present study are summarized in Table 4.4.

Therefore, in this present case study author compared the microbial communities of activated sludge from large and small scale wastewater treatment plants and introduced a workflow for analysis of large data sets and together with graphical interpretation.

Material and Methods

Nearly 100,000 sequences (Table 4.4) were obtained form 20 activated sludge samples and analyzed by process explained in Fig. 4.13.

Table 4.4 Sample details including no. of quality sequences, Mode of treatment, Period of sampling and Treatment plant statistics

	WWTP ID	No.of Sequences	Treatment Mode	Period of Sampling	Serviced Population	Serviced area (ha)
Small Scale	SA	3301	Oxidation Ditch	Oct. 2007	~ 1500	~ 65
	SB	3245	Oxidation Ditch	Oct. 2007	~ 20,000	~ 1000
	SC	2776	Oxidation Ditch	Oct. 2007	~ 3500	~ 150
Large Scale	LA1	6494	Conventional	July 2007	~ 800,000	
	LA2	7930	Anaerobic Aerobic	Feb. 2010		
	LA3	5007	Anaerobic Aerobic	Feb. 2010		
	LB1	4815	Pre-denitrification Process with coagulant addition	Nov. 2009	~ 1,325,000	~ 15,000
	LB2	5247	Pre-denitrification Process with coagulant addition	Nov. 2009	~ 165,000	
	LC1	5934	Conventional	May 2009	~ 260,000	
	LC2	6258	Anaerobic Aerobic	May 2009		
	LD1	4490	Conventional	Nov. 2007	~ 1,200,000	~ 15,000
	LD2	3352	Conventional	Aug. 2008		
	LE	5526	Conventional	May 2009	~ 2,100,000	
	LF	3839	Conventional	Sep. 2009	~ 80,000	~ 1500
	LG	4269	Conventional	July 2009	~ 500,000	~ 6500
	LH	4048	Anaerobic Aerobic	Jan. 2010	~ 82,000	~ 1000
	LI1	6451	Pre-denitrification Process with coagulant addition	Nov. 2009	~ 225,000	~ 6500
	LI2	5215	Pre-denitrification Process with coagulant addition	Nov. 2009		
	LI3	4097	Anaerobic Anoxic Oxic	Oct. 2008		
	LI4	5066	Bardenpho	Oct. 2008		

The methodology shown in Fig. 4.13 comprised of three main sections,

1. Selection of quality reads by QIIME

   The original set of sequences were analysed by QIIME to select the quality reads with length longer than 200bp. The sequences were matched with barcode and reads were separated into each sample. The details of sample names, barcode and primer sequences were given in mapping file (text file).

2. QIIME data analysis

   Sequences resulted from quality analysis were further analysed and microbial communities were grouped to their OTU level. Analysis was performed by running QIIME (Quantitative Insight Into Microbial Ecology, Caporaso et al., 2010) commands on terminal application.

   - Operational Taxonomic Units (OTUs) based community identification (97% similarity)
   - Pick representative sequences
   - Taxonomic assignment by RDP classifier)
   - Construction of phylogenetic tree by fastree method

   At the end of analysis, statistics on microbial community analysis were given in text based file.

3. OTUMAMi data interpretation

   Analysis results were imported to a pyrosequencing workflow programme, OTUMAMi. It calculated the developed individual read counts and cumulative counts per sample. The counts are given as fraction for total sequences in individual samples. The layout of the module was re-arranged to interpret the microbial community composition for each sample and highlighted with colors (2 dimensional heatmap).

In addition,

- Microbial compositions were compared at Phylum, Class and Family level.
- Partial phylogenetic tree was constructed for representative sequences form major OTUs merged to Family level and combined with heatmap.

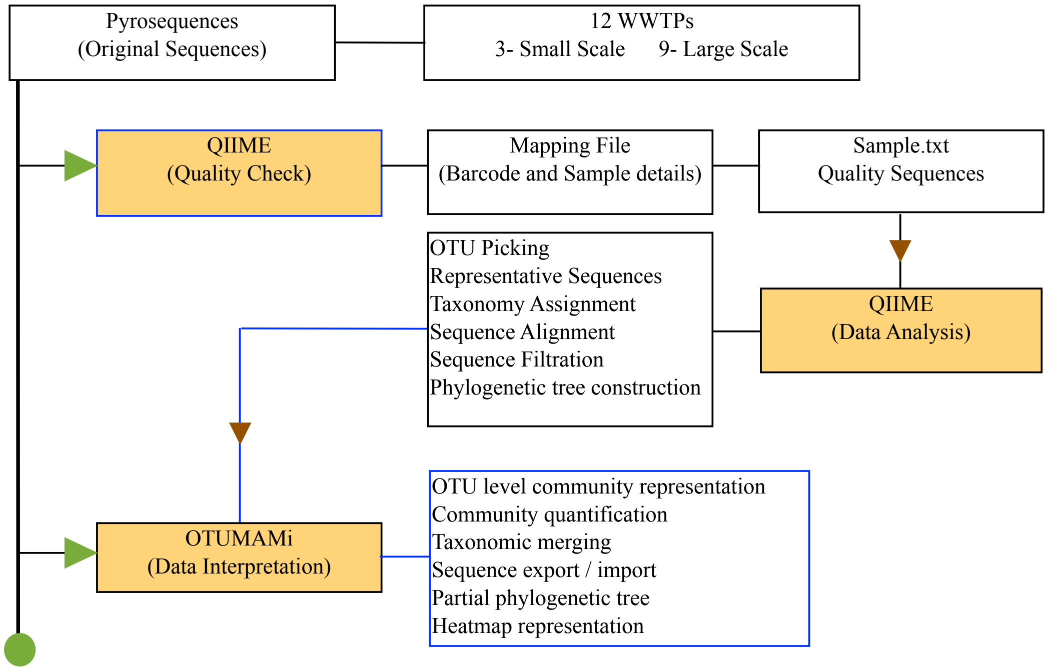


Fig.4.13 Workflow for case study 3

Results and Discussion

1. QIIME data analysis

The sequence analysis was performed using QIIME commands. The outcomes of each analysis generate text files with sequences statistics and following section explained the nature of text based data.

Initial file name: CS3_20_27f.txt

```
>SA_27f_WW_117
GATGAACGCTAGCGGCAGGCCTAATACATGCAAGTCGTGGGGCAGCAGGTGTA........
>SA_27f_WW_139
AGCGAACGTTTGCGGCGGGCCTAACACATGCAAGTCGAACGGGTTGGCAACAA........
>SA_27f_WW_256
AGCGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGGCGTAGCCAAT........
>SA_27f_WW_296
AACGAACGCTGACGGCGGGGCTTAGGCATGCAAGTTGAGCGAGAAAGCCGCAA........
>SA_27f_WW_318
GATTAACGCTAGCGGCAGGCCTAATACATGCAAGTCGAACGGGATTATTGGTAG........
```

Sequences were arranged in FASTA format (>) and saved in a text file. Then, following command lines were executed in "Terminal programme" (Command line execution for Mac OSX).

```
macqiime pick_otus.py -i CS3_20_27f.txt -m uclust -s 0.97; macqiime pick_rep_set.py -i uclust_picked_otus/CS3_20_27f_otus.txt -f CS3_20_27f.txt; macqiime assign_taxonomy.py -i CS3_20_27f.txt_rep_set.fasta; macqiime align_seqs.py -i CS3_20_27f.txt_rep_set.fasta -t /qiime2/core_set_aligned.fasta.imputed.txt -m pynast; macqiime filter_alignment.py -i pynast_aligned/CS3_20_27f.txt_rep_set_aligned.fasta -m /qiime2/lanemask_in_1s_and_0s.txt; macqiime make_phylogeny.py -i CS3_20_27f
```

- OTU Picking (File name: CS3_20_27f_otus.txt)

Sequences were grouped based on 97% similarity.

```
0 A3_27f_WW_9844 A2_27f_WW_22348 A3_27f_WW_13934
1 A2_27f_WW_20517 A2_27f_WW_16674
2 A3_27f_WW_6176
3 AIAS_27f_35700 C2_27f_WW_3532 F1_27f_WW_9586
4 AKAS_27f_8045 AIAS_27f_35393 AKAS_27f_4074 AKAS_27f_12371
```

OTU numbers indicated from 0 to last number and sequence IDs were assigned to each OTU number as shown above.

- Representative Sequences
  (File name: CS3_20_27f.txt_rep_set.fasta)

representative sequnce ID was selected from each OTU based on their similarity indices.

```
>0 A3_27f_WW_9844
GATGAACGCTAGCGGCAGGCCTAACACATGCAAGAGAG...
>1 A2_27f_WW_20517
AGTGAATGCTTGAGTATGCTTTACACATGCAAGTCTAAA...
>10 AKAS_27f_8049
ATTGAACGCTGGCGGCGTGCTTTACACATGCAAGTCGAGC...
>100 B2_27f_WW_8619
AACGAACGTTAGCGGCGCGCTTAACACATGCAAGTCGAGC...
```

- Taxonomy Assignement

  (File Name: CS3_20_27f.txt_rep_set_tax_assignments.txt)

The representative sequnces were presnted with their taxonomic affliation and percentage of assigenement.

```
788 AIAS_27f_37984 Root;Bacteria 0.970
10607 C1_27f_WW_16056 Root;Bacteria 1.000
7654 C2_27f_WW_2942   Root;Bacteria;Proteobacteria;Betaproteobacteria 1.000
7475 AKAS_27f_1291 Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales 0.890
```

Taxonomy aasignemnt were performed for 85% taxonomy matching and similarity indices were also given along with the OTU number and ID.

- Sequence Alignment

  (File Name: CS3_20_27f.txt_rep_set_aligned.txt)

Representative sequnces were aligned as shown in FASTA aligned file. And, by prefiltering, poorly aligned sequences were removed. And, rest given in, CS3_20_27f.txt_rep_set_aligned_pfiltered.fasta file.

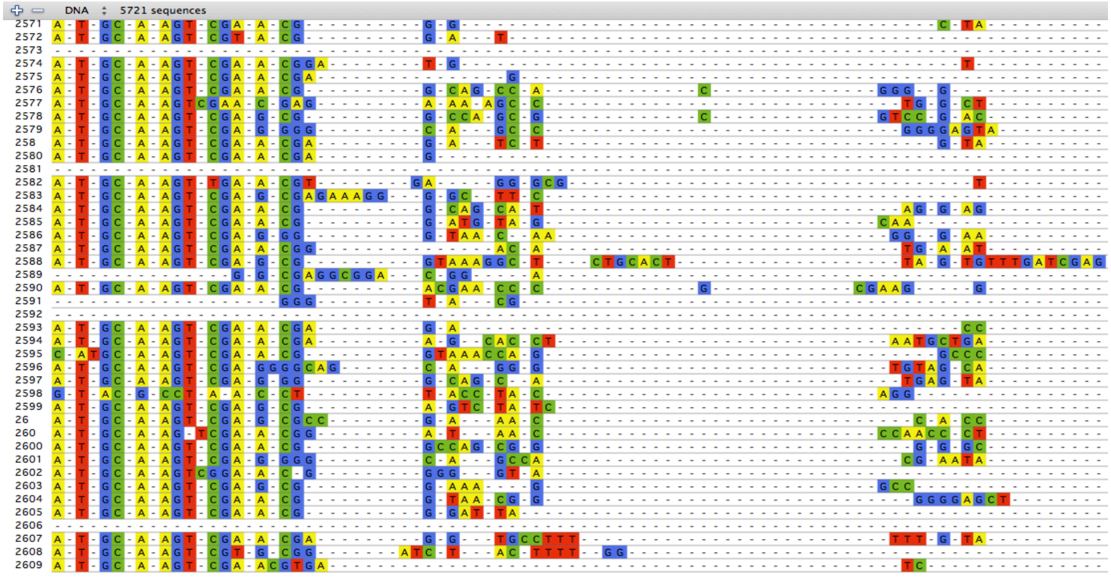


Fig.4.14 Representative sequence alignment

- Phylogenetic tree file
  (File name: CS3_20_27f.txt_rep_set_aligned_pfiltered.tre)

```
('1614_Root_Bacteria_Proteobacteria_Betaproteobacteria_Rhodocyclales_Rhodocyclaceae':0.00015,'5
18_Root_Bacteria_Proteobacteria_Betaproteobacteria':0.0303)0.888:0.00409,('1895_Root_Bacteria_Pr
oteobacteria_Betaproteobacteria_Rhodocyclales_Rhodocyclaceae':0.00958,'3854_Root_Bacteria_Prote
obacteria_Betaproteobacteria_Rhodocyclales_Rhodocyclaceae':0.03371)0.878:0.00015,('4268_Root_B
acteria_Proteobacteria_Betaproteobacteria_Rhodocyclales_Rhodocyclaceae_Zoogloea':0.00015
```

Tree file contains, OTU number, taxonomy and tree order.

For comparison of microbial communities at small and large scale treatment plants, Following data files were selected and imported to OTUMAMi.

- CS3_20_27f.txt
- CS3_20_27f_otus.txt
- CS3_20_27f.txt_rep_set_tax_assignments.txt
- CS3_20_27f.txt_rep_set_aligned_pfiltered.fasta
- CS3_20_27f.txt_rep_set_aligned_pfiltered.tre

The QIIME platform helped to analyse nearly 100,000 reads with in short period time and generate many sequence statistic data for further analysis and interpretaion.

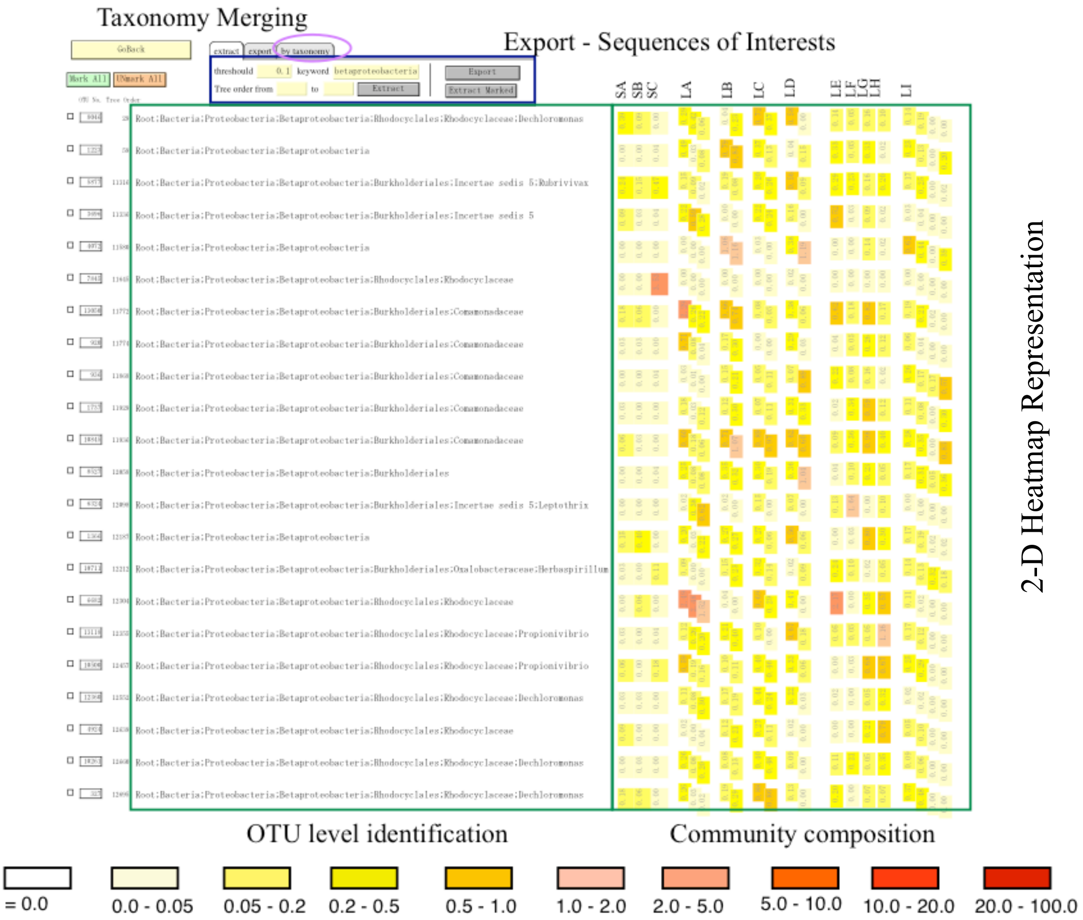
2. OTUMAMi computation

In fig. 4.15 presents the microbial community compositions for each activated sludge sample. The layout illustrated OTU level identities and their quantity. The application of colors helped to examine the microbial population in small and large treatment plants. And, also samples collected at different mode of operation in same wastewater treatment plant also compared.

Therefore OTUMAMi layout effectively present following features,

1. Easy layout for 1-D or 2-D heatmap representation allowing comparison of sample dynamics
2. Compositions were computed at their OTU level.
3. Extract sequences of interest, setting Keyword and Threshold values

4. Export selected sequences as text file for further analysis (Fig.4.17)
5. Entry for taxonomic merging from Phylum to Genus level (Fig.4.16)
6. Construction of phylogenetic trees with OTU level taxonomy identity and combine with heatmap for better resolution (Fig.4.18)

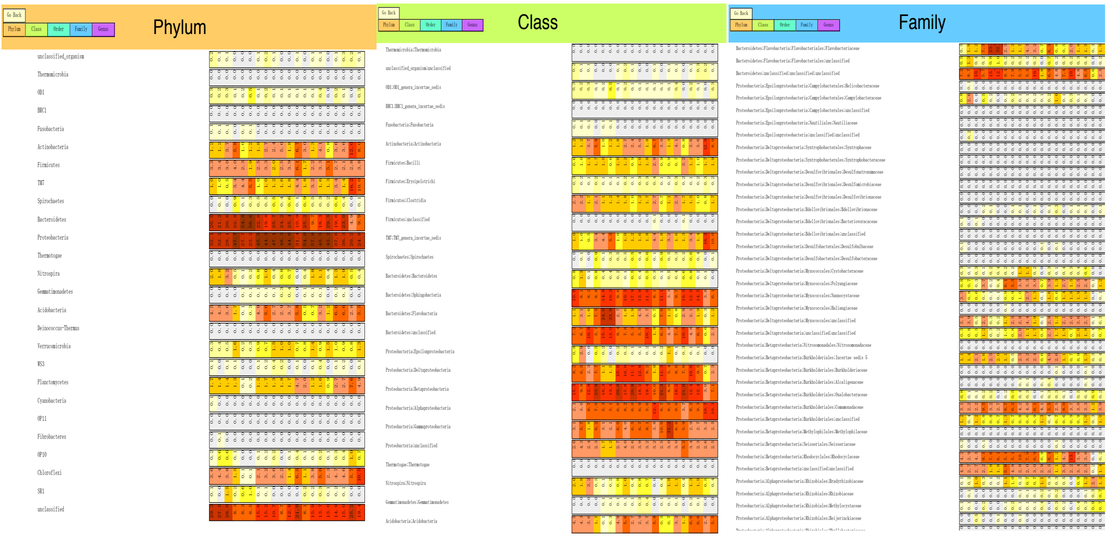


threshould 0.1 keyword betaproteobacteria
Tree order from to

□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Dechloromonas
□ Root;Bacteria;Proteobacteria;Betaproteobacteria
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Incertae sedis 5;Rubrivivax
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Incertae sedis 5
□ Root;Bacteria;Proteobacteria;Betaproteobacteria
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Incertae sedis 5;Leptothrix
□ Root;Bacteria;Proteobacteria;Betaproteobacteria
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Propionivibrio
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Propionivibrio
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Dechloromonas
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Dechloromonas
□ Root;Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Dechloromonas

Fig. 4.15 OTU representation and community composition for sequence reads

As explained above, reads were exported and merged in to their Phylum and Class level. Reads counts major Phyla and Classes were used to construct charts using MS Excel and explained as in Fig. 4.17.



Phylum

unclassified_organism
Thermomicrobia
OD1
BRC1
Fusobacteria
Actinobacteria
Firmicutes
TM7
Spirochaetes
Bacteroidetes
Proteobacteria
Thermotogae
Nitrospira
Gemmatimonadetes
Acidobacteria
Deinococcus-Thermus
Verrucomicrobia
WS3
Planctomycetes
Cyanobacteria
OP11
Fibrobacteres
OP10
Chloroflexi
SR1
unclassified

Class

Thermomicrobia;Thermomicrobia
unclassified_organism;unclassified
OD1;OD1_genera_incertae_sedis
BRC1;BRC1_genera_incertae_sedis
Fusobacteria;Fusobacteria
Actinobacteria;Actinobacteria
Firmicutes;Bacilli
Firmicutes;Erysipelotrichi
Firmicutes;Clostridia
Firmicutes;unclassified
TM7;TM7_genera_incertae_sedis
Spirochaetes;Spirochaetes
Bacteroidetes;Bacteroidetes
Bacteroidetes;Sphingobacteria
Bacteroidetes;Flavobacteria
Bacteroidetes;unclassified
Proteobacteria;Epsilonproteobacteria
Proteobacteria;Deltaproteobacteria
Proteobacteria;Betaproteobacteria
Proteobacteria;Alphaproteobacteria
Proteobacteria;Gammaproteobacteria
Proteobacteria;unclassified
Thermotogae;Thermotogae
Nitrospira;Nitrospira
Gemmatimonadetes;Gemmatimonadetes
Acidobacteria;Acidobacteria

Family

Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae
Bacteroidetes;Flavobacteria;Flavobacteriales;unclassified
Bacteroidetes;unclassified;unclassified;unclassified
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Helicobacteraceae
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae
Proteobacteria;Epsilonproteobacteria;Campylobacterales;unclassified
Proteobacteria;Epsilonproteobacteria;Nautiliales;Nautiliaceae
Proteobacteria;Epsilonproteobacteria;unclassified;unclassified
Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae
Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfonatronumaceae
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae
Proteobacteria;Deltaproteobacteria;Bdellovibrionales;Bdellovibrionaceae
Proteobacteria;Deltaproteobacteria;Bdellovibrionales;Bacteriovoracaceae
Proteobacteria;Deltaproteobacteria;Bdellovibrionales;unclassified
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae
Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae
Proteobacteria;Deltaproteobacteria;Myxococcales;Cystobacteraceae
Proteobacteria;Deltaproteobacteria;Myxococcales;Polyangiaceae
Proteobacteria;Deltaproteobacteria;Myxococcales;Nannocystaceae
Proteobacteria;Deltaproteobacteria;Myxococcales;Haliangiaceae
Proteobacteria;Deltaproteobacteria;Myxococcales;unclassified
Proteobacteria;Deltaproteobacteria;unclassified;unclassified
Proteobacteria;Betaproteobacteria;Nitrosomonadales;Nitrosomonadaceae
Proteobacteria;Betaproteobacteria;Burkholderiales;Incertae sedis 5
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae
Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae
Proteobacteria;Betaproteobacteria;Burkholderiales;unclassified
Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae
Proteobacteria;Betaproteobacteria;unclassified;unclassified
Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae
Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylocystaceae
Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae

Fig.4.16 Taxonomic Assignments were merged at Phylum, Class and Genus level

Microbial communities identified at OTU level were merged into Phylum, Class and Family level. The community compositions were calculated and color intensities were sample. The unclassified microbial composition also calculated here.

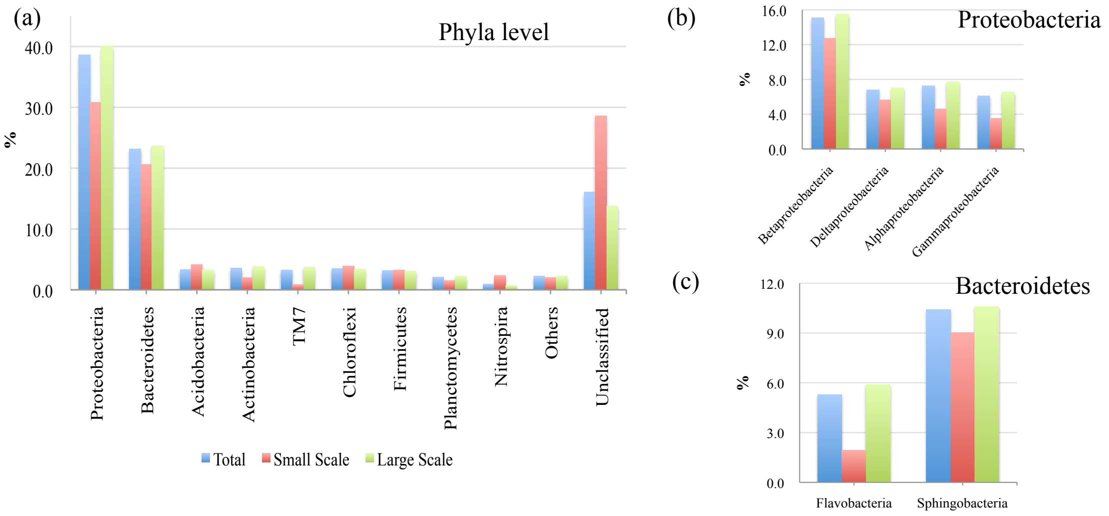


Fig. 4.17 Composition of the reads from samples. (a) Phylum level (b) Class level in Phylum Proteobacteria (c) Classes level in Phylum Bacteroidetes

Further, species level lineage difference was found in Phylum Actinobacteria. Different lineages found in small scale plants and large scale plants were identified by their distribution and phylogenetic information as shown in Fig. 4.18.

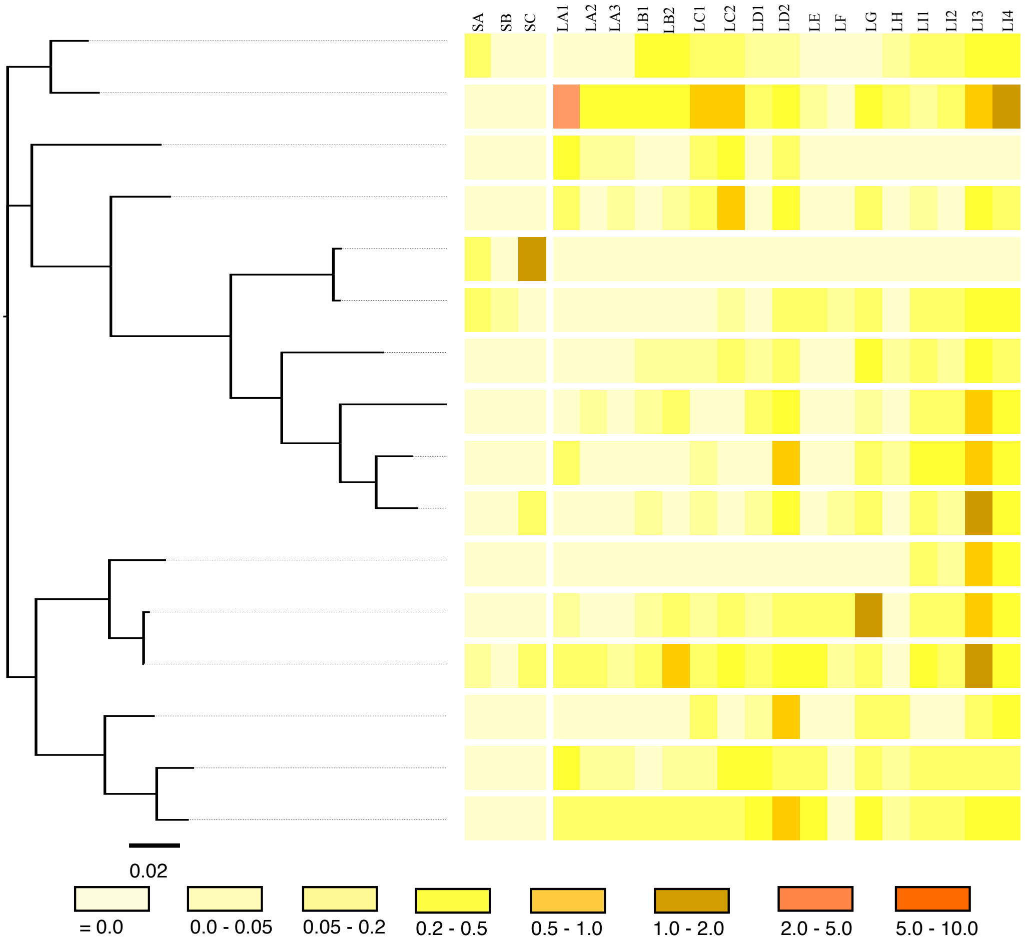


Fig.4.18 Partial Phylogenetic tree and Heatmap representation for OTU level identities of Phylum Actinobacteria

In the present study, the microbial community compositions were compared at different levels of phylogeny. The distribution of microbial community structures were clearly presented by OTUMAMi and by using the developed workflow data analysis became faster.

And, the proposed methodology also can manage larger number of reads (approximately 100,000) effectively, which is one important achievement for the management of pyrosequencing data.

Limitations and Recommendations

QIIME as a Data Analysis Tool,

1. The QIIME process recovered more quality sequences than RDP pyrosequencing pipeline. The matching of barcode cannot be customized in RDP and it search for exact barcode sequence by default. In QIIME, barcode matching is controlled. And, sequences were labeled with sample name and barcode + primer sequence pair was removed.

2. In microbial classification, QIIME process allows processing of massive number of sequence without any limitations. The processing time is only couple of hours depend on the number of sequences. Further, sequences from different samples can easily concatenated and process in single run.

3. QIIME provides a single platform for data analysis. Many tools are combined and while executing commands user can continue to do the analysis without importing or exporting data files manually. At each step, output file automatically saved in to folder given by the user. Text based files are produced (aaaa.txt) at each analysis step can open in any text editor software.

OTUMAMi for Data Interpretation

1. Computational tool was developed to manage text-based results and convert the data into graphical representation. As shown in Fig. 4.14, high-resolution data obtained in pyrosequencing data analysis were effectively presented by OTUMAMi tool.

2. Arrangement of layout can customized by user in order to present the microbial community dynamics based on sample type, climatic condition, time f sampling or mode of treatment.

3. The database functionalities given with the tools helps the user to search, export of sequences of interests. And the exported data files can further analyze by external softwares like MS Excel or tree building tools. And also, data file can used to perform QIIME analysis and obtained more detailed information.

4.6 Discussion

The case studies explained in present chapter were present the development of the workflow pyrosequencing data analysis.

The first case study only for limited number of reads where manual data management were mostly applied using many data management tools. Then, in second study, database tools and computational approaches were introduced. However, much time spent on the data analysis and the calculations on community compositions were done by computational method. The microbial community structures and their compositions were effectively presented using relational database application (heatmap).

The workflow introduced in the third study,

> Effectively manage large number of pyrosequence data much faster and microbial community structures were presented with their community compositions at a higher resolution.

The methodologies applied in all three studies were compared and summarized in Table 4.5

Therefore, the developed workflow introduced as a efficient computational tool for pyrosequence data analysis and interpretation.

Table 4.5 Comparison on data analysis methodologies applied in all three case studies

Main Characteristics	Study 1	Study 2	Study 3
Number of Samples	2	8	20
Number of Sequences	4,495	~ 13,000	~ 100,000
Sequence Quality Analysis	RDP	RDP	QIIME
Data Analysis - 1. Classification	RDP	RDP	QIIME
2. Quantification	Manual	Manual & Computational	Computational
Data Interpretation	Ms Excel Charts	Excel Charts & Computational module	Computational Module (OTUMAMi)
Phylogenetic Analysis	ARB	ARB	Full & Partial Trees By QIIME
Analysis Time	5 - 7 days (Depends on number of samples)	5 - 7 days (Depends on number of samples)	5 - 6 hours (Depends on number of sequences)

Chapter 5

Development of Floc Library

5.1 Introduction

Flocs of activated sludge are composed of different microorganisms, and they are diversified with different shapes, sizes and colors. There have been a couple of studies to clarify details of flocs in activated sludge, as listed below.

1. Impact of structure characteristics on sludge floc stability (Wilen, Jin and Lant, 2003)
2. Characterization of activated sludge flocs by confocal laser scanning microscopy and image analysis (Schmid et al., 2003)
3. Three-Dimensional Modeling of an Activated Sludge Floc (Zartarian et al., 1997)
4. Microbial community structure in activated sludge floc analyzed by fluorescence in situ hybridization and its relation to floc stability (Wilen et al., 2008)

Yet, information to connect the morphological and microbial characteristics of activated sludge flocs is scarce. The goal of this study was to develop a data organization tool, or a library, to study morphological and microbial data for flocs in activated sludge. Here, to characterize microbial communities in flocs, Restriction Fragment Length Polymorphism (RFLP) analysis was employed because of the easiness to obtain data. Also, it will help to co-relate microbial community structures with different forms of flocs.

The library comprised of two layouts. First layout shows the general characteristics of activated sludge flocs and second to show the differences in microbial communities related to each individual floc. The library was developed on FileMaker pro Advanced (Version.10.0) software.

5.2 Materials and methods

5.2.1 Sampling, floc isolation and storage

Activated sludge samples were collected during the aerobic phase from four (A-D) laboratory scale Sequencing Batch Reactors (SBR) at different time intervals. These reactors had been operated under sequencing anaerobic and aerobic conditions. As the scope of the study was to develop a data organizing tool, the details of the operational conditions and the performances of these reactors are not described here. Brief description of the reactors are in the Appendix III. Just after sampling, sludge samples were observed under microscope. The sludge characteristics (Fig.5.1) were observed based on nature of sludge, floc shape, size and color.

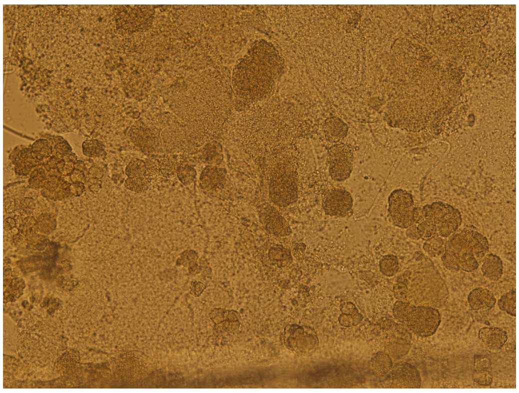


Fig.5.1 Activated sludge (x100)

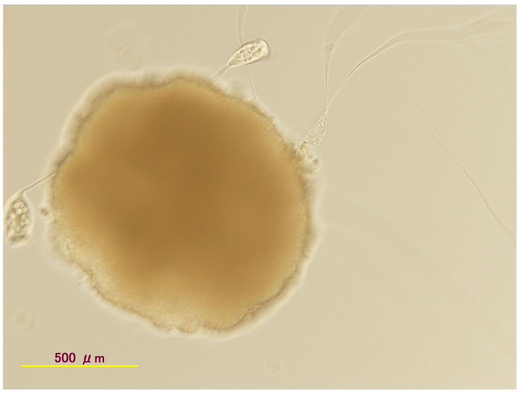


Fig.5.2 Isolated floc (x100)

To analyze microbial population in each floc, activated sludge mixed liquor was diluted, dispensed on glass plate as 10μL aliquots, and single flocs were collected from aliquots containing only one floc in each. Each of the collected floc was transferred to 1.5mL plastic tube containing 1ml of Milli-Q (Fig.5.2). Initially flocs in Milli-Q were stored at -20°C and later samples were transferred to 50% ethanol medium and stored under -80°C. In total, nearly 100 flocs were isolated from time to time sampling. Each sample was labeled by date of sampling and name of reactor (Appendix III). During microscopic examination, microphotographs were taken with a x10 magnification objective lens.

5.2.2 DNA extraction and PCR amplification

Individual floc samples were sonicated by a 250DA Advanced Digital Sonifier (Branson) with a special micro tip at an amplitude of 30% (20W) for 20 seconds. The sonified liquid was directly used as the template DNA for Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) for DNA amplification using PrimeScript® One Step RT-PCR Kit Ver.2 (Takara) and a Thermal Cycler Dice (Takara, Japan). Part of sonified liquid for each of the samples were stored with 50% ethanol at -80°C and rest kept under -80°C.

The content of the RT-PCR reaction mixture was as follows (per sample),

2X buffer	10.0µl
DNAase Free H2O	5.2µl
RNAase	0.4µl
PSE	0.8µl
Primer (Forward)*	0.8µl
Primer (Reverse)*	0.8µl
Sample	2.0µl
Total Volume	**20.0 µl**

PSE – PrimeScript 1 step Enzyme Mix containing
PrimeScript RTase
DNA polymerase
RNase Inhibitor

27 forward and 519 reverse primers were used.

27f – AGAGTTTGATCMTGGCTCAG

519r – GWATTACCGCGGCKGCTG

Thermal cycle for RT-PCR ,

50°C	|94°C	|94°C	| 55.3°C	|72°C	| 72°C	| 4°C
1800 sec	120 sec	30 sec	30 sec	30 sec	600sec	store
		|	30 cycles			

The PCR product concentrations were determined by PicoGreen (Quant-iT^{TM} PicoGreen ® dsDNA Reagent and Kits) from Invitrogen according to manufacture's instructions. The PCR products were subjected to the RFLP analysis, which is explained in the next section.

5.2.3 Restriction Fragment Length Polymorphism (RFLP) Data

Initially, RFLP analysis was performed using *Rsa*I (GT|AC) as the restriction enzyme for DNA digestion. Due to limited number of restriction sites, the number of fragments was found to be less. Therefore, later in the analysis, HhaI (GCG|C) restriction enzyme was used for digestion.

The digestion reaction was performed as follows,

Reaction mixture per sample,

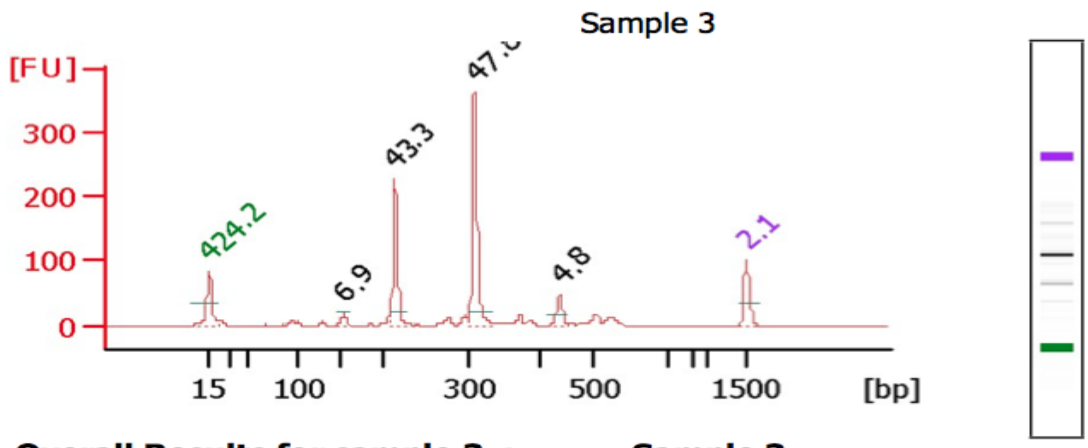
10X buffer	1.00µl
*Hha*I	0.24µl
BSA	1.00µl
Sterile Milli-Q	2.66µl
Sample	5.0µl
Total Volume	**10.0 µl**

BSA – Bovine Serum Albumin

Reaction was performed under, 37°C for 4 hours and 65°C for 15 minutes using Thermal Cycler Dice (Takara, Japan).

Digested samples were analyzed and RFLP profiles were obtained by using an Agilent BioAnalyzer 2100 with a DNA 1000 Assay KIT. Obtained profiles were examined for their different restriction fragment lengths (bp) and their intensities (Fig.5.3). The gel electrophorogram obtained for selected 12 samples were given in Fig. 5.4 as an example.

Finally, 72 samples were selected, microphotographs and RFLP profiles for were obtained, and were used as the data to develop the floc library.



Overall Results for sample 3 : **Sample 3**

Number of peaks found: 4

Peak table for sample 3 : **Sample 3**

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	15	4.20	424.2	Lower Marker
2		154	0.70	6.9	
3		215	6.15	43.3	
4		309	9.75	47.8	
5		435	1.38	4.8	
6	▶	1,500	2.10	2.1	Upper Marker

Fig.5.3 RFLP profile generated from Agilent DNA 2100 Assay

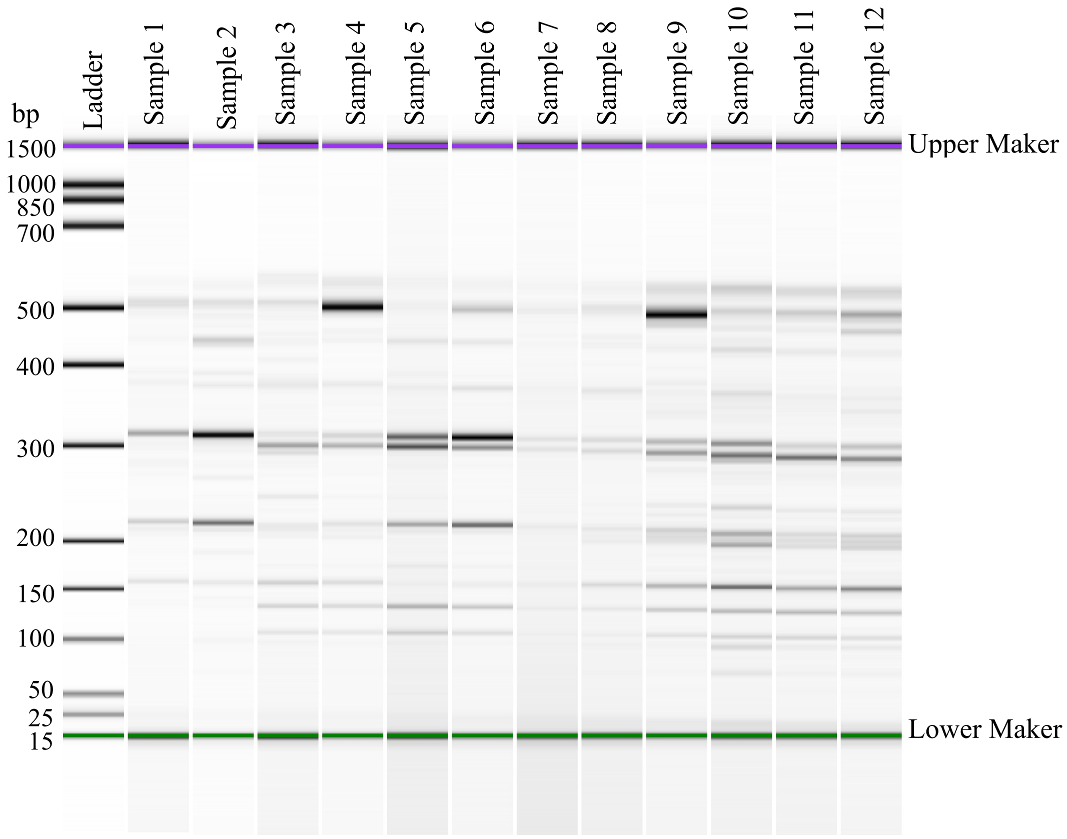


Fig.5.4 Gel image generated from Agilent DNA 2100 Assay

5.3 Development of Floc Library

The floc library was developed as a collection of 72 activated sludge flocs (S1-S72) with different, shapes, sizes and colors. RFLP patterns and correspondent restriction fragment lengths were also included in the library as a measure of differences in microbial communities. The complete sets of data were systematically imported to a library module developed by FileMaker Pro Advanced (version 10.0) software.

5.3.1 Structure of the database

The floc library database developed here is mostly like a card-type database without relationship. The database has following fields.

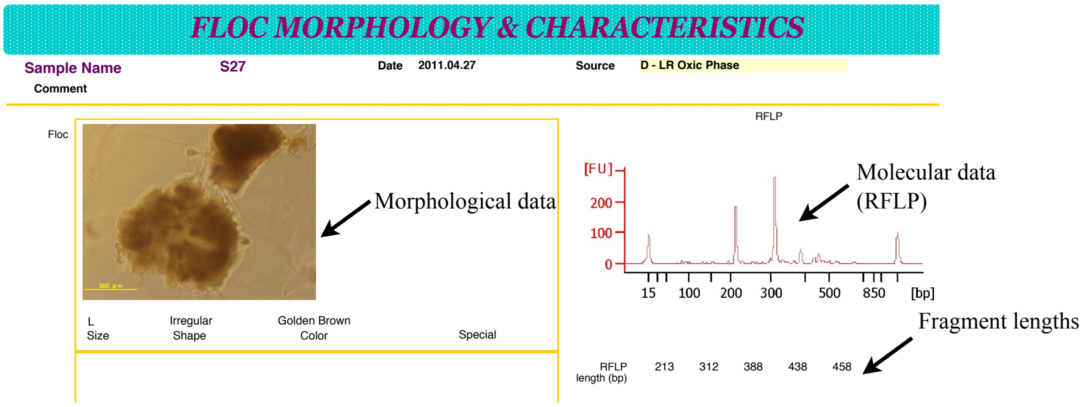
- Fields to describe source of sample
  - ➢ date of sampling
  - ➢ source reactor name
- Fields to describe characteristics of flocs
  - ➢ Microphotograph image
  - ➢ Floc size
  - ➢ Floc shape
  - ➢ Floc color
  - ➢ Special characteristics such as Zoogloeal, tetrad, or filamentous
- Fields to describe microbial population
  - ➢ RFLP profile image
  - ➢ Fragments sizes (12 fields)

5.3.2 Interface to enter data

The floc library has two layouts. One is to enter data on morphological and microbial information of flocs (Fig.5.5). And another is a gallery of flocs with different morphology to help users choose proper words to describe the flocs (Fig. 5.6).

Grouping was done based on the criteria described in “Microscopic Examination of the Activated Sludge process by Gerardi, 2008”, and to help classification, the gallery as shown in Fig. 5.6 was introduced.

On the left side of Fig. 5.5, fields related to the morphological information of the floc are located, including the field to store microphotograph of the floc. On the right side of Fig. 5.5, fields for molecular information (RFLP results) are located. Users can manually enter these fields. Or, users can use a spreadsheet file to enter data, and then import to the floc library database.



FLOC MORPHOLOGY & CHARACTERISTICS

Sample Name S27 Date 2011.04.27 Source D - LR Oxic Phase

Comment

Floc

L	Irregular	Golden Brown	
Size	Shape	Color	Special

RFLP

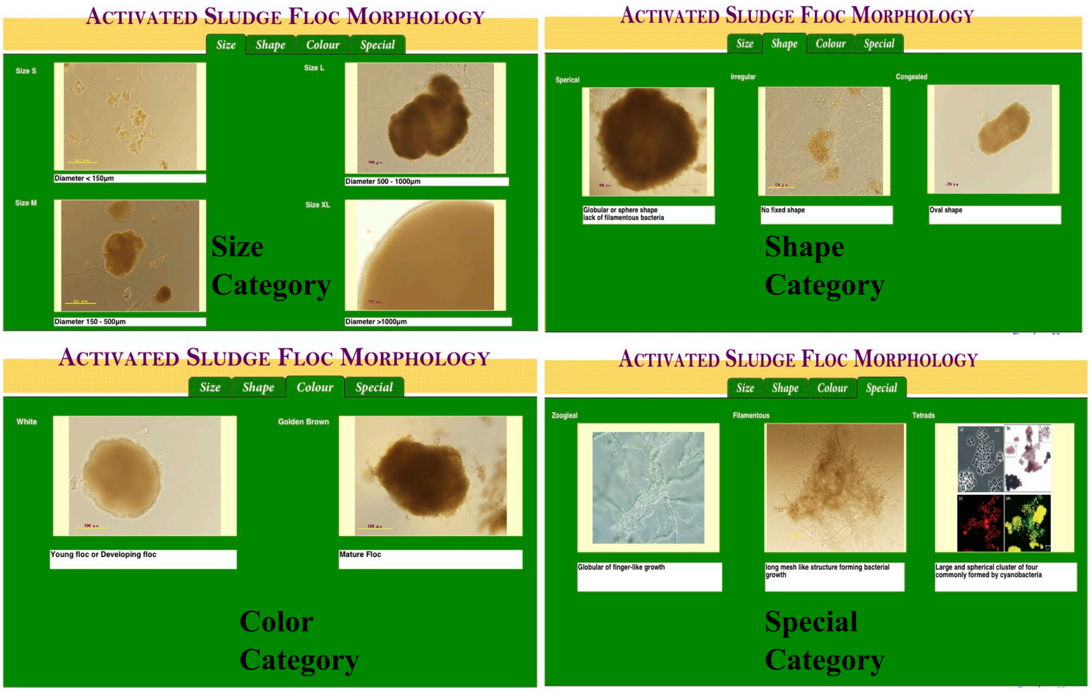
RFLP length (bp)	213	312	388	438	458

Morphological data

Molecular data (RFLP)

Fragment lengths

Fig. 5.5 Floc library record showing morphological and molecular data



ACTIVATED SLUDGE FLOC MORPHOLOGY

Size Shape Colour Special

Size S: Diameter < 150µm

Size L: Diameter 500 - 1000µm

Size M: Diameter 150 - 500µm

Size XL: Diameter >1000µm

Size Category

ACTIVATED SLUDGE FLOC MORPHOLOGY

Size Shape Colour Special

Sperical: Globular or sphere shape lack of filamentous bacteria

Irregular: No fixed shape

Congealed: Oval shape

Shape Category

ACTIVATED SLUDGE FLOC MORPHOLOGY

Size Shape Colour Special

White: Young floc or Developing floc

Golden Brown: Mature Floc

Color Category

ACTIVATED SLUDGE FLOC MORPHOLOGY

Size Shape Colour Special

Zoogleal: Globular of finger-like growth

Filamentous: long mesh like structure forming bacterial growth

Tetrads: Large and spherical cluster of four commonly formed by cyanobacteria

Special Category

Fig. 5.6 Gallery view and grouping criteria for activated sludge floc

5.4 Application of floc library data

The comparison I and II were performed to present utility of floc library as a interface to compare molecular data and morphological data. Three records were listed below.

Comparison I

Sample: S08 & S07

Date of Sampling: same date

Source: same laboratory reactor

- Morphology

  In Fig. 5.7 records labeled as (a), sample S08 and (b), sample S07 shared similar floc characteristics except the size of floc. Both flocs were collected from filamentous sludge sampled on same day.

- RFLP data

  S08 and S07 shared 150bp and 286bp fragment sizes where, 488bp length only found in S08 and 127bp size belongs to S07.

The comparison between morphology and microbial data showed, the differences in morphology effect on the microbial community structures found. In morphology S07 floc large in size and can categorized as a mature floc compared to S08. Therefore, depends on the floc constituents (organic matter, extracellular substances and secondary metabolites) and internal (chemical) environment, attached microbial population may get different.

Comparison II

Sample: S07 & S31

Date of Sampling: different

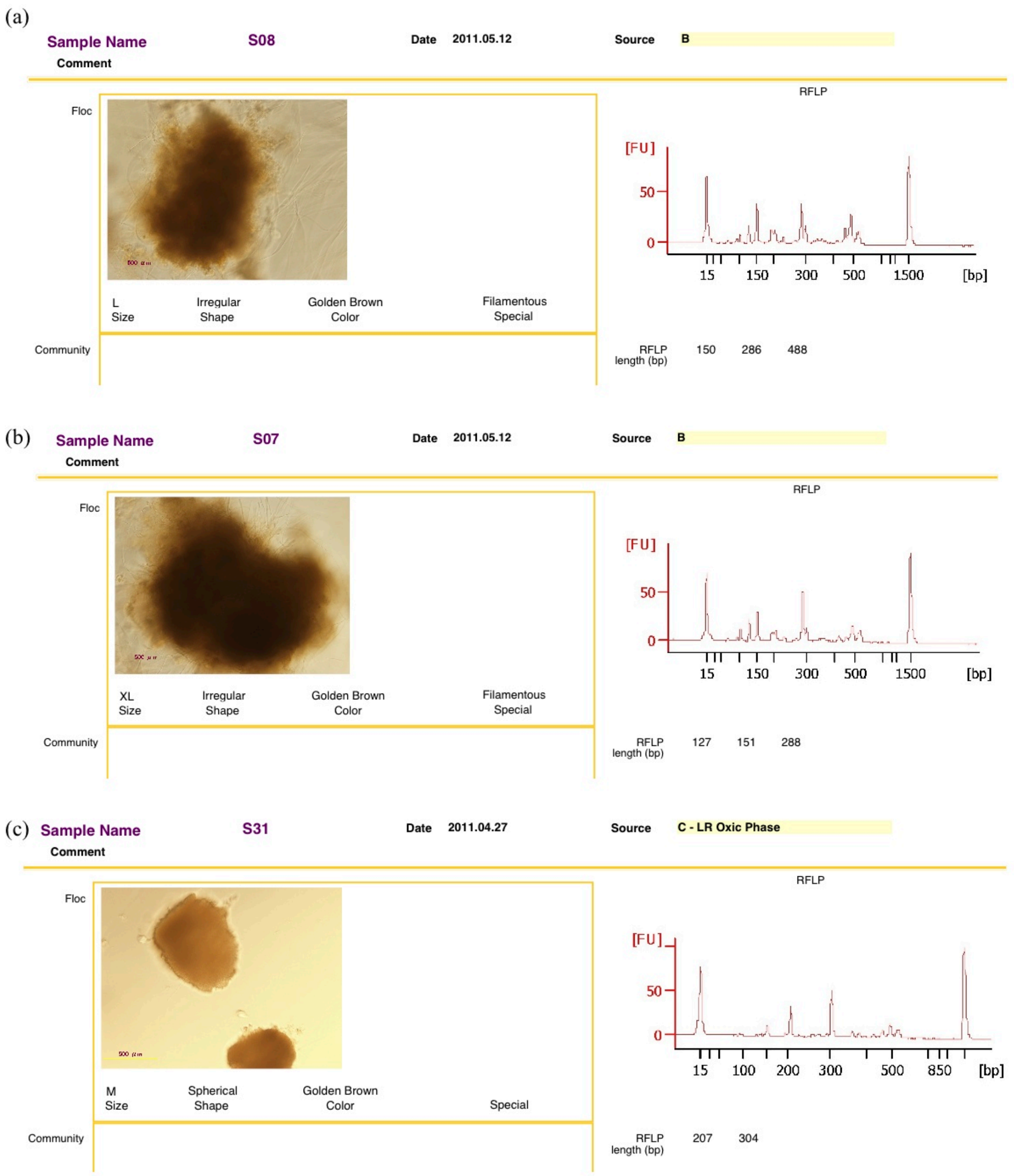
Source: different

- Morphology

  In Fig. 5.7 records labeled as (b), sample S07 and (c), sample S31 characterized with different morphological features related to size, shape, filamentous nature and source of sludge.

- RFLP data

  S07 and S31 produced totally different peak patterns. Sample S07 gave 127bp, 151bp and 288bp fragment sizes where, sample S08 produced 207bp and 304bp.



(a)

Sample Name S08 Date 2011.05.12 Source B

Comment

Floc

L Size · Irregular Shape · Golden Brown Color · Filamentous Special

Community

RFLP length (bp) 150 286 488

(b)

Sample Name S07 Date 2011.05.12 Source B

Comment

Floc

XL Size · Irregular Shape · Golden Brown Color · Filamentous Special

Community

RFLP length (bp) 127 151 288

(c)

Sample Name S31 Date 2011.04.27 Source C - LR Oxic Phase

Comment

Floc

M Size · Spherical Shape · Golden Brown Color · Special

Community

RFLP length (bp) 207 304

Fig. 5.7 Records from floc library (a) and (b) Filamentous activated sludge floc isolated from same source (c) Spherical floc isolated from different activated sludge source.

As mentioned in comparison I, it's confirmed that morphology itself effect on the microbial community structures or the organization of different microbial clusters give rise to different floc morphologies. And, also comparison II showed there are many types of microbial communities (Floc- formers) found in a single floc structure.

Therefore, the interface given with the floc library helps the user to find a similar floc structures for a unknown activated sludge sample. Thereby, user can predict the molecular data information associated and understand the diversity of microbial communities without applying advanced analysis tools.

Thus, as explained in the present study, the developed floc library provides a support to store morphological and molecular information. Further, it acts as an interface for providing information to user and helps comparison studies.

5.5 Discussion

5.5.1 Difficulties in obtaining data

Floc Isolation:

Isolation of single floc structures through microscopic observations was a difficulty found in the study. Different methods were practiced to select single flocs from activated sludge samples. Preparation of floc smears, sample dilution methods and sludge settling techniques were applied and those were randomly used on different sludge samples depends on the nature of the sludge (bulky sludge, pin flocs).

In the present study, morphological characters were explained by using microphotographs. Therefore, manual handling of floc is important to obtain high-resolution snapshot with a clear view of floc without any damage. Therefore, collection of higher number of flocs was essential since they can loose their stability during sample handling.

Failures at PCR:

The general size of a floc can range from 100μm – 1000μm. It's found that A few flocs size smaller than 100μm tends to fail at PCR. The less amount of template DNA is the main cause and therefore RT-PCR (Reverse Transcription PCR) was introduced. The RNA is the target template in RT-PCR. In a bacterial cell the number of copies of RNA is always higher than that for DNA. And, also due to the instability of RNA and digestion by RNAase RT-PCR may also failed. However, a few sets of samples were removed only 72 left with successful results in RT-PCR.

5.5.2 Possible Improvements for the floc library

As, a recommendation present library can be further developed with the addition of many more flocs structures and then users can compare the sludge morphology and microbial communities without performing molecular analyses like T-RFLP or sequencing.

As a database tool, the structure of the present library can be further improved. In the present structure, card-type table view is used. With the addition of many records tables can arranged into relation database structure, where user can select and present records of floc based on it's source, size or shape.

And, also it's important to minimize manual data feeding on floc morphology and RFLP data. As a option single table can defined (main table) to hold all the record as a data repository.

Chapter 6

Development of Reference Database

As was presented in Chapter 4, an efficient workflow for pyrosequencing data handling was developed. But in order to compare more samples, it should be useful to have a collection of important bacterial species found in activated sludge samples. The “reference database” was developed for this purpose. While there already are DNA databases such as GenBank, DDBJ and EMBL, these are not specifically for activated sludge microorganisms. The reference database to be developed here will be the one specialized on activated sludge microorganisms.

6.1 Identification of major microbial communities

Pyrosequence results obtained for 64 different samples (described in Chapter 3 and in Appendix I) were utilized and the data was analyzed by QIIME and organized by OTUMAMi module. To utilize reversely-read reads, reads starting with 519r primer regions were converted into its reverse complement sequence by executing perl command as below.

```
open(IN, "<Original_519r.txt");
@file = <IN>;
$thisfile = @file[0];
close(IN);
$connected="";
foreach $txt(@file)
{
($accession, $sequence)=($txt=~ /^(.*)\t(.*)$/);
$sequence = reverse $sequence;
$sequence =~ tr/ATGCatgc/TACGtacg/;
$addition = $accession . "\t" . $sequence . "\n";
$connected = $connected . $addition;
};
open(OUT,">Original_519rev.txt");
print(OUT $connected);
close(OUT);
```

Here,

0riginal_519r.txt : File containing 519r reads split after quality checking (Input)

0riginal_519rev.txt: File name including reverse complement reads of 519r sequences (0utput)

Extracting Major Communities

The results of complete data analysis for samples produced nearly 41,000 OTUs. Their community compositions were computed and displayed in sample basis. In the present study, major microbial communities were extracted by setting the threshold value, the minimum value for the average value of the percentage of the OTU for all samples, to 0.008.

The Partial sequences, taxonomic identities and OTU numbers for these OTUs were exported and analyzed by RDP classifier. RDP classifier analysis provided extended taxonomic assignments for each OTU with the confidence percentage of assignment. The outcomes used for the development of the reference database model will be presented in section 6.2.

6.2 Development of Reference Database

6.2.1 Structure of the Reference Database

The reference database has four tables as shown in Fig. 6.1. The "my sequence" table is for original OTU database imported from OTUMAMi. The reference database should be prepared only for supposed to be important OTUs. And thus, only important OTU data imported to "my sequence" table needs to be selected. And the selected OTU data are saved in "Reference Database". The table "Sequence Select" is used to help selecting important OTU data from the data in "my sequence". And finally, table "temporal" is to help finalizing table "Reference Database".

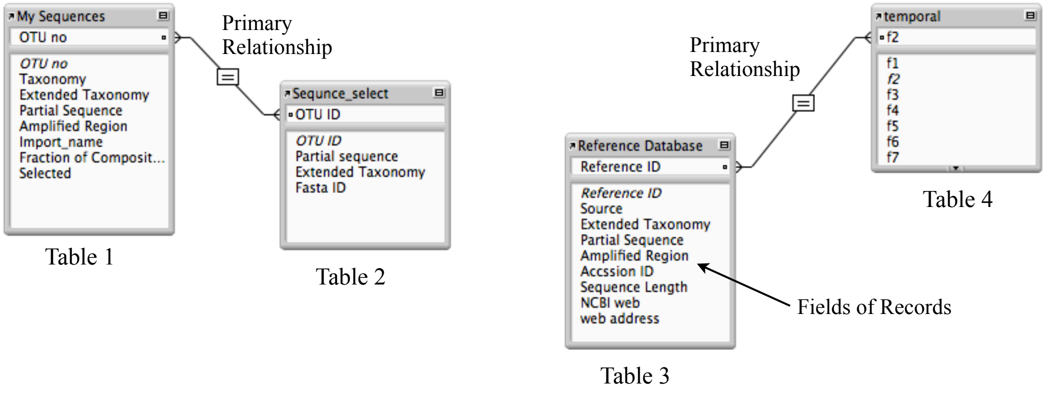


Fig. 6.1 Structure of reference database showing data tables and relationships

6.2.2 Outline of the work

The way to use the Reference Database is as below.

1) Import original OTU data to table "my sequences"

2) Process original OTU data by RDP Classifier on the Internet to obtain detailed description of the taxonomic assignment result.

3) Import the result from 2) to table "my sequence" while using Reference ID (Serial number plus OTU ID) number as the key. Then, detailed taxonomic information will be added to table "my sequence" for each OTU.

4) Select OTUs from the records in table "my sequence".

   Here, the selected OTU data is sent to table "Sequence Select" automatically. The OTUs in table "my sequence" have already been selected from the whole OTUs stored in OTUMAMi. Additional selection step here is to reduce the number of sequences to send to RDP Sequence Match, which allows only up to 2000 sequences per batch.

5) The selected OTUs data accumulated in table "Sequence Select" are exported as a text file in FASTA format.

6) The exported FASTA format file is analyzed by RDP Sequence Match, which can be found at RDP database homepage (Cole et al., 2009). A text file containing the OTU numbers and the accession number of their nearest match DNA in GenBank is returned.

7) The text file returned from RDP Sequence Match is imported to the table "temporal".

8) In table "Reference Database", the field for accession number is still empty. The accession number is searched from the table "temporal" by using Reference ID number as the key.

9) Table "Reference Database" is displayed on layout "Reference Database", where detailed information on the closest match sequence found for the OTU is retrieved and displayed to help identifying the nature of the microorganism related to the OTU.

6.2.3 Details of the layouts

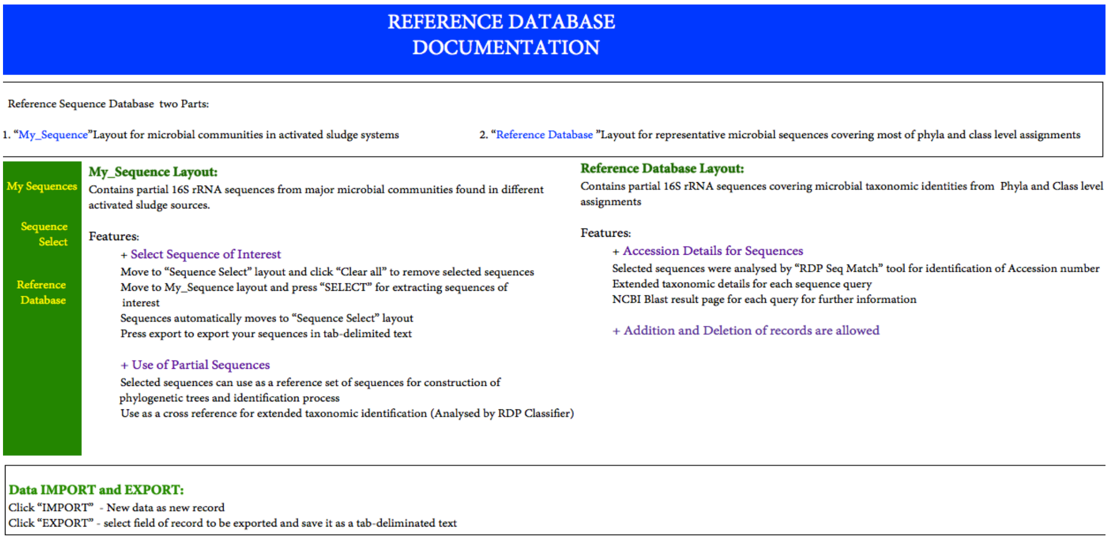
The database has four layouts as below.

- Home: a portal page
- My sequences: to show and work on data in table "my sequences"
- Sequence select: to show selected data from table "my sequences". Data shown in this layout is stored in table "Sequence select".
- Reference database: to show data in table "Reference database"

The details of the four layouts are as follows.

Home Page

Brief description on layouts in this database, navigation guidelines, and sequence handling methods are explained in the home page. This will be documentation on the use of database (Fig.6.2).



REFERENCE DATABASE
DOCUMENTATION

Reference Sequence Database two Parts:

1. "My_Sequence"Layout for microbial communities in activated sludge systems

2. "Reference Database "Layout for representative microbial sequences covering most of phyla and class level assignments

My_Sequence Layout:
Contains partial 16S rRNA sequences from major microbial communities found in different activated sludge sources.

Features:

+ Select Sequence of Interest
Move to "Sequence Select" layout and click "Clear all" to remove selected sequences
Move to My_Sequence layout and press "SELECT" for extracting sequences of interest
Sequences automatically moves to "Sequence Select" layout
Press export to export your sequences in tab-delimited text

+ Use of Partial Sequences
Selected sequences can use as a reference set of sequences for construction of phylogenetic trees and identification process
Use as a cross reference for extended taxonomic identification (Analysed by RDP Classifier)

Reference Database Layout:
Contains partial 16S rRNA sequences covering microbial taxonomic identities from Phyla and Class level assignments

Features:

+ Accession Details for Sequences
Selected sequences were analysed by "RDP Seq Match" tool for identification of Accession number
Extended taxonomic details for each sequence query
NCBI Blast result page for each query for further information

+ Addition and Deletion of records are allowed

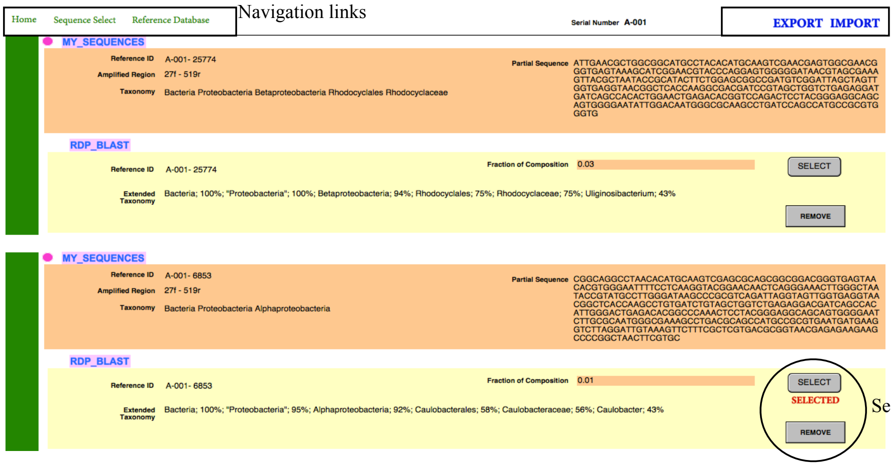
Data IMPORT and EXPORT:
Click "IMPORT" - New data as new record
Click "EXPORT" - select field of record to be exported and save it as a tab-deliminated text

Fig.6.2 Documentation on Sequence data management and Layout Navigation

My Sequences

"My Sequence" layout contains the details of major microbial communities selected from complete sequence data set which is saved in OTUMAMi. Their partial sequence, taxonomic identity in short format, taxonomic identity in detailed format, average composition of community and OTU number (Reference ID) are shown in this layout (Fig.6.3).

Here, further to reduce the number of OTUs, users can select OTUs of interest. By clicking "SELECT" button, information on the OTUs will be copied to table "Sequence Select".



Navigation links

Export / Import recorc

Selected record

Serial Number A-001

Reference ID A-001- 25774
Amplified Region 27f - 519r
Taxonomy Bacteria Proteobacteria Betaproteobacteria Rhodocyclales Rhodocyclaceae
Partial Sequence ATTGAACGCTGGCGGCATGCCTACACATGCAAGTCGAACGAGTGGCGAACGGGTGAGTAAAGCATCGGAACGTACCCAGGAGTGGGGGATAACGTAGCGAAAGTTACGCTAATACCGCATACTTCTGGAGCGGCCGATGTCGGATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGGGTG

Reference ID A-001- 25774
Fraction of Composition 0.03
Extended Taxonomy Bacteria; 100%; "Proteobacteria"; 100%; Betaproteobacteria; 94%; Rhodocyclales; 75%; Rhodocyclaceae; 75%; Uliginosibacterium; 43%

Reference ID A-001- 6853
Amplified Region 27f - 519r
Taxonomy Bacteria Proteobacteria Alphaproteobacteria
Partial Sequence CGGCAGGCCTAACACATGCAAGTCGAGCGCAGCGGCGGACGGGTGAGTAACACGTGGGAATTTTCCTCAAGGTACGGAACAACTCAGGGAAACTTGGGCTAATACCGTATGCCTTGGGATAAGCCCGCGTCAGATTAGGTAGTTGGTGAGGTAACGGCTCACCAAGCCTGTGATCTGTAGCTGGTCTGAGAGGACGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAGTTCTTTCGCTCGTGACGCGGTAACGAGAGAAGAAGCCCCGGCTAACTTCGTGC

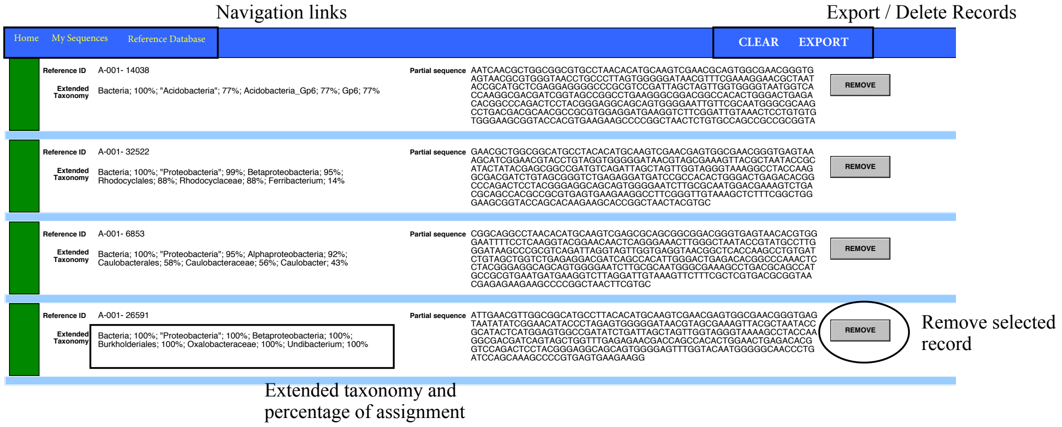
Reference ID A-001- 6853
Fraction of Composition 0.01
Extended Taxonomy Bacteria; 100%; "Proteobacteria"; 95%; Alphaproteobacteria; 92%; Caulobacterales; 58%; Caulobacteraceae; 56%; Caulobacter; 43%

Fig. 6.3 Detailed view of two sequence records in "My Sequence" layout.

Sequence Select

This layout contains the OTU information selected in the layout "My database"(Fig. 6.4). Before use, a user needs to refresh the records to delete current records by clicking the button "CLEAR". The "EXPORT " button exports the data into a tab-delimitated text file. In exporting, a user can select the type of data and it's arrangement.

Export and re-analyse the data by RDP "SeqMatch" tool, and best matching sequences and their accession numbers are obtained. The obtained results come in a tab-delimited text file.



Reference ID A-001- 14038

Extended Taxonomy Bacteria; 100%; "Acidobacteria"; 77%; Acidobacteria_Gp6; 77%; Gp6; 77%

Partial sequence
```
AATCAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAACGCAGTGGCGAACGGGTG
AGTAACGCGTGGGTAACCTGCCCTTAGTGGGGGATAACGTTTCGAAAGGAACGCTAAT
ACCGCATGCTCGAGGAGGGGCCCGCGTCCGATTAGCTAGTTGGTGGGGTAATGGTCA
CCAAGGCGACGATCGGTAGCCGGCCTGAAGGGCGGACGGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTGTTCGCAATGGGCGCAAG
CCTGACGACGCAACGCCGCGTGGAGGATGAAGGTCTTCGGATTGTAAACTCCTGTGTG
TGGGAAGCGGTACCACGTGAAGAAGCCCCGGCTAACTCTGTGCCAGCCGCCGCGGTA
```

Reference ID A-001- 32522

Extended Taxonomy Bacteria; 100%; "Proteobacteria"; 99%; Betaproteobacteria; 95%; Rhodocyclales; 88%; Rhodocyclaceae; 88%; Ferribacterium; 14%

Partial sequence
```
GAACGCTGGCGGCATGCCTACACATGCAAGTCGAACGAGTGGCGAACGGGTGAGTAA
AGCATCGGAACGTACCTGTAGGTGGGGGATAACGTAGCGAAAGTTACGCTAATACCGC
ATACTATACGAGCGGCCGATGTCAGATTAGCTAGTTGGTAGGGTAAAGGCCTACCAAG
GCGACGATCTGTAGCGGGTCTGAGAGGATGATCCGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGACGAAAGTCTGA
CGCAGCCACGCCGCGTGAGTGAAGAAGGCCTTCGGGTTGTAAAGCTCTTTCGGCTGG
GAAGCGGTACCAGCACAAGAAGCACCGGCTAACTACGTGC
```

Reference ID A-001- 6853

Extended Taxonomy Bacteria; 100%; "Proteobacteria"; 95%; Alphaproteobacteria; 92%; Caulobacterales; 58%; Caulobacteraceae; 56%; Caulobacter; 43%

Partial sequence
```
CGGCAGGCCTAACACATGCAAGTCGAGCGCAGCGGCGGACGGGTGAGTAACACGTGG
GAATTTTCCTCAAGGTACGGAACAACTCAGGGAAACTTGGGCTAATACCGTATGCCTTG
GGATAAGCCCGCGTCAGATTAGGTAGTTGGTGAGGTAACGGCTCACCAAGCCTGTGAT
CTGTAGCTGGTCTGAGAGGACGATCAGCCACATTGGGACTGAGACACGGCCCAAACTC
CTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCCAT
GCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAGTTCTTTCGCTCGTGACGCGGTAA
CGAGAGAAGAAGCCCCGGCTAACTTCGTGC
```

Reference ID A-001- 26591

Extended Taxonomy Bacteria; 100%; "Proteobacteria"; 100%; Betaproteobacteria; 100%; Burkholderiales; 100%; Oxalobacteraceae; 100%; Undibacterium; 100%

Partial sequence
```
ATTGAACGTTGGCGGCATGCCTTACACATGCAAGTCGAACGAGTGGCGAACGGGTGAG
TAATATATCGGAACATACCCTAGAGTGGGGGATAACGTAGCGAAAGTTACGCTAATACC
GCATACTCATGGAGTGGCCGATATCTGATTAGCTAGTTGGTAGGGTAAAAGCCTACCAA
GGCGACGATCAGTAGCTGGTTTGAGAGAACGACCAGCCACACTGGAACTGAGACACG
GTCCAGACTCCTACGGGAGGCAGCAGTGGGGAGTTTGGTACAATGGGGGCAACCCTG
ATCCAGCAAAGCCCCGTGAGTGAAGAAGG
```

Fig. 6.4 Records from” Sequence Select”: Selected sequences from “My sequences”

Reference Database

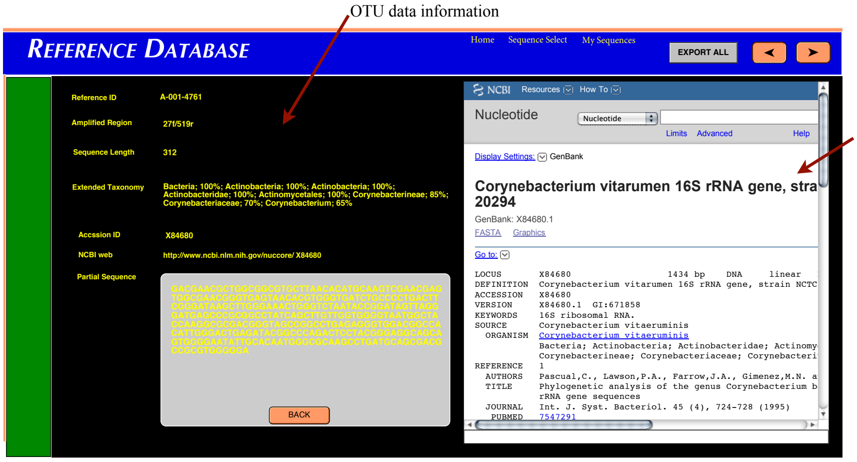
The ”Reference Database” layout shows OTU data that have been selected as their significance and are used as “reference” to analyze microbial communities from different activated sludge samples. The data shown in “Reference Database” are stored in table “Reference Database”.

And the data stored here are expected to be used as the reference to compare and characterize microbial populations in unknown samples from activated sludge processes.

To help users to grasp the characteristics of the OTUs, available information of the closest match species is presented in this layout. That is, accession number for the closest match sequence in the GenBank database is imported to table “Reference Database”, and is utilized to retrieve information in NCBI GenBank (Fig. 6.5). To show the detailed information on the closest match species, WebViewer function in FileMaker Pro is utilized.

The present layout present the representative reference sequences with Genbank annotation details for with information of closest matching closet species.

OTU data information



Reference ID	A-001-4761
Amplified Region	27f/519r
Sequence Length	312
Extended Taxonomy	Bacteria; 100%; Actinobacteria; 100%; Actinobacteria; 100%; Actinobacteridae; 100%; Actinomycetales; 100%; Corynebacterineae; 85%; Corynebacteriaceae; 70%; Corynebacterium; 65%
Accssion ID	X84680
NCBI web	http://www.ncbi.nlm.nih.gov/nuccore/ X84680
Partial Sequence	GACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGAGTGGCGAACGGGTGAGTAACACGTGGGTGATCTGCCCCTGACTTCGGGATAAGCTTGGGAAACTGGGTCTAATACCGGATAGTTAGGGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCTACCAAGGCGCGACGGGTAGCCGGCCTGAGAGGGTGGACGGCCACATTGGGACTGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGGGGGA

Corynebacterium vitarumen 16S rRNA gene, stra
20294

GenBank: X84680.1

```
LOCUS       X84680                  1434 bp    DNA     linear
DEFINITION  Corynebacterium vitarumen 16S rRNA gene, strain NCTC
ACCESSION   X84680
VERSION     X84680.1  GI:671858
KEYWORDS    16S ribosomal RNA.
SOURCE      Corynebacterium vitaeruminis
  ORGANISM  Corynebacterium vitaeruminis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomy
            Corynebacterineae; Corynebacteriaceae; Corynebacteri
REFERENCE   1
  AUTHORS   Pascual,C., Lawson,P.A., Farrow,J.A., Gimenez,M.N. a
  TITLE     Phylogenetic analysis of the genus Corynebacterium b
            rRNA gene sequences
  JOURNAL   Int. J. Syst. Bacteriol. 45 (4), 724-728 (1995)
   PUBMED   7547291
```

NCBI's GenBank Annotation

-Accession ID
- Species level identification
  (Type cultures)
- Isolation source
- Full or partial sequences
  (FASTA or GenBank Format)

Fig.6.5 Reference database record for major OTU level lineage

Present database provides detailed information for the selected major microbial communities in activated sludge samples. It will be a reference to sequence data analysis and construction of phylogenetic trees with unknown sequence data.

6.3 Discussion

The developed database not only store sequence information of important microbial groups in activated sludge but also it can further incorporate more detailed classification information. In each step more and more information is added to the original partial sequence.

The major OTU level sequences were given with a reference ID that formulated by serial number and OTU number. This considered as a unique field for the identification of the species.

The present database developed as a collection of OTUs found at major quantities and use it's sequence and related data as a reference data for classification of unknown set of sequences. Therefore, it 's framework can further developed to select and add minor OTUs found in activated sludge samples or OTUs with higher abundance in different samples as reference sequnces.

Chapter 7

Conclusion and Recommendation

The present study was performed for the development of database tools to manage data obtained on microbial population from wastewater treatment processes.

As explained in chapter 4, 5 and 6 three different database tools were introduced for proper management of microbial population data like 16S rRNA sequences obtained by pyrosequencing and Restriction Fragment Length Polymorphism (RFLP). The database tools developed are as follows.

➢ Workflow for pyrosequencing data

The proposed workflow for management of 16S rRNA pyrosequencing data were applied and tested on data obtained from different activated sludge samples.

- The case studies discussed in chapter 4, presented the development of methodology from manual data management to computational approach. The benefits of computational approach, for fastening data analysis and storage of sequence information were also confirmed.

- As explained in Chapter 4, several methodologies were applied and finally computational workflow developed (OTIMAMi) was recommended. The use of massive number of pyrosequences for data analysis was also tested on the computational workflow.

Recommendation: Present workflow can further evolve to make link with external bioinformatics tools like phylogenetic tree drawing or statistical tools for calculating correlation values for selected communities.

- Development of Floc library

The floc library tool was developed to aid the organization of morphological data and molecular data of sludge flocs. The developed tool enables, easier comparison of morphological and molecular data for different flocs. To help describing morphological characteristics, a gallery page is included in the database.

- Better interface for present the relationship between morphological and molecular data
- Gallery view for activated sludge flocs with different shapes, sizes from different activated sludge sources.

Recommendation: The framework to be further improved to mange more new data sets like, community identities by 16S rRNA pyrosequencing. It helps to compare molecular data in two methods (Pyrosequencing and RFLP).
Also, the library data will be a ideal source wastewater treatment plants, for comparing activated sludge morphology and determine its microbial community compositions with out applying advanced methodologies.

- Development of Reference Database

The reference database was developed as a collection of OTUs found in higher quantity and at different levels of taxonomic assignments to serve as the “reference” OTUs.

- The database provides an interface for add extra taxonomic information (Extended taxonomy, GenBank annotation) and an interface to add new OTUs to reference OTUs.

Recommendation: The database further developed to allow user to select sequences of from different taxa, minor OTUs or abundant OTUs found in different samples. Also, it can link with OTUMAMi and quantitative information on selected OTUs can then present in the reference database.

LIST OF REFERENCES

Journal Articles & Proceedings

Blackwood, C.B, Marsh, T., Kim, S. and Paul, E.A. (2003). Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Applied and Environmental Microbiology*, **69**(2), 926.

Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., et al., with Yadhukumar. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, **32**(4), 1363-71.

Caffaro-Filho, R.A., Fantinatti-Garboggini, F. and Durrant, L.R., (2007). Quantitative analysis of Terminal Restriction Fragment Length Polymorphism (T-RFLP) microbial community profiles: peak height data showed to be more reproducible than peak area. *Brazilian Journal of Microbiology*, **38**, 736–738.

Cannone J.J., Subramanian S., Schnare M.N., Collett J.R., D'Souza L.M., Du Y., Feng B., Lin N., Madabusi L.V., MÜller K.M., Pande N., Shang Z., Yu N., and Gutell R.R. (2002). The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other RNAs. *BioMed Central Bioinformatics*, **3**:15.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Hurtley, G. A., Kelley, S., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., Mcdonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). Correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nature*, **7,** 335-336.

Cole J.R., Chai B., Marsh T.L, Farris R.J., Wang Q., Kulam S.A., Chandra S., McGarrell D.M., Schmidt T.M., Garrity G.M. and Tiedje J.M. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research*, **31**(1), 442-443.

Dickie, I. A and FitzJohn, R.G. (2007). Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review. Mycorrhiza, **17**(4), 259-70.

Dixon, M. T. and Hills, D. M. (1993). Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution*, **10**(1), 256-267.

Eikelboom, D. (1975). Filamentous organisms observed in activated sludge. *Water Research*, **9**(4), 365-388.

García Martín, H., Ivanova, N., Kunin, V.,Warnecke, F.,Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A. A., Szeto, E., Dalin, E., Putnam, N. H., Shapiro, H. J., Pangilinan, J. L., Rigoutsos, I., Kyrpides, N. C., Blackall, L. L., McMahon, K. D and Hugenholtz, P. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology*, **24**(10), 1263-1269.

Hamady, M. and Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, 19(7), 1141-1152.

Head, I., Saunders, J. and Pickup, R. (1998). Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microbial Ecology*, **35**(1), 1-21.

Hodkinson, B.P., and Lutzoni, F. (2009). A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. Symbiosis **49**: 163-180.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A and Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*, **55**(5), 856-66.

Kent, A.D., Smith, D. J., Benson, B.J. and Triplett, E.W. (2003). Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. *Applied and Environmental Microbiology*, **69**(11), 6768.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D., and Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*, **35**(18), e120.

Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J., and Woese, C. R. (1997). The RDP (Ribosomal Database Project). Nucleic Acids Research**, 25**(1), 109-11.

Marsh, T. L., Saxman, P., Cole, J. and Tiedje, J. (2000). Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Applied and Environmental Microbiology*, **66**(8), 3616-3620.

Mino, T. (2000). Microbial selection of polyphosphate-accumulating bacteria in activated sludge wastewater treatment processes for enhanced biological phosphate removal. *Biochemistry. Biokhimii͡a*, **65**(3), 341-8.

Murat E.A., Ferris, M. J. and Taylor, C. M. (2011). A framework for analysis of metagenomic sequencing data. *Pacific Symposium on Biocomputing,* 131-41.

Nübel, U., Garcia-Pichel, F., and Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, **63**, 3327–3332.

Nyrén, P. (2007). "The History of Pyrosequencing". *Methods Molecular Biology* **373**, 1–14. PMID 17185753.

O'Brien, H., Miadlikowska, J., and Lutzoni, F. 2005. Assessing host specialization in symbiotic Cyanobacteria associated with four closely related species of the lichen fungus *Peltigera*. *European Journal of Phycology*, **40**, 363-378

Pandey, J., Ganesan, K. and Jain, R.K. (2007). Variations in T-RFLP profiles with differing chemistries of fluorescent dyes used for labelling the PCR primers. *Journal of microbiological methods*, **68**(3), 633-638.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**(21), 7188-96.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G. and Triplett, E.W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, **1**(4), 283-90.

Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research*, 3-11.

Ronaghi, M., and Elahi, E. (2002). Pyrosequencing for microbial typing. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, **782**(1-2), 67-72.

Rudi, K., Skulberg, O.M., and Jakobsen, K.S. (1998). Evolution of cyanobacteria by exchange of genetic material among phyletically related strains. *Journal of Bacteriology* **180**, 3453–3461.

Rudi, K., Skulberg, O.M., Larsen, F., and Jakobsen, K.S. (1997). Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7 and V8. *Applied and Environmental Microbiology* **63**, 2593-2599.

Satoh, H., Ogawa, A. and Mino, T. (2009). Effect of activated sludge extract on microbial population in activated sludge screened by incubation on microplates. Proceedings of Environmental Research Forum **46,** 503-510

Schmid, M., Thill, A., Purkhold, U., Walcher, M., Bottero, J.Y., Ginestet, P., Nielsen, P. H., Wuertz, S. and Wagner, M. (2003). Characterization of activated sludge flocs by confocal laser scanning microscopy and image analysis. *Water Research*, **37(**9), 2043-2052.

Shendure, J. and Ji, H., (2008). Next-generation DNA sequencing. *Nature Biotechnology*, **26**(10), 1135-45.

Slater, F. R., Johnson, C. R., Blackall, L. L., Beiko, R. G. and Bond, P. L. (2010). Monitoring associations between clade-level variation, overall community structure and ecosystem function in enhanced biological phosphorus removal (EBPR) systems using terminal-restriction fragment length polymorphism (T-RFLP). *Water Research*, **44**(17, 4908-4923.

Turner, S., Pryer, K.M., Miao, V.P.W., and Palmer, J.D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. Journal of Eukaryotic Microbiology **46**: 327–338.

Unno, T., Jang, J., Han, D., Kim, J. H., Sadowsky, M. J., Kim, O. Sun., Chun, J., Hur, H. G., (2010). Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environmental science & technology*, **44(**20), pp.7777-82.

Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., & Daims, H. (2002). Microbial community composition and function in wastewater treatment plants. *Antonie van Leeuwenhoek*, **81**(1-4), 665-80.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, **73**(16), 5261-5267.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* 173: 697-703.

Wilen, B., (2003). Impacts of structural characteristics on activated sludge floc stability. *Water Research*, **37**(15), 3632-3645.

Zartarian, F., Mustin, C., Villemin, G., Thill, A., Bottero, J. Y., (1997). Three-Dimensional Modelling of an Activated Sludge Floc. *Society*, **30**(11), 187-192.

Book References

Lane, D.J. 1991. 16S/23S rRNA sequencing. In: Nucleic acid techniques in bacterial systematics. Stackebrandt, E., and Goodfellow, M., eds., John Wiley and Sons, New York, NY, 115-175.

Marco, D. 2010. Metagenomics: Theory, Methods and Applications. Caister Academic Press. ISBN 978-1-904455-54-7

Beychok M. R. 1967. Aqueous Wastes from Petroleum and Petrochemical Plants (1st ed.). John Wiley & Sons Ltd. New York, NY.

Madigan, M.T., Martinko, J.M., Dunlap, P.V and Clark, D.P. 2009. Brock Biology of Microorganisms (12th ed.) Pearson Education Inc. ISBN 978-0-13-232460-1

Bitton, G. 2005. Wastewater Microbiology (3rd ed.). John Wiley and Sons, New Jersey.

Geradi, M.H. 2008. Microscopic Examination of the Activated Sludge Process. John Wiley and Sons, New York, NY. 105-120.

Prosser, S. and Gripman, S. 2010. FileMaker Pro 11: The Missing Manual. O'Reilly Media Inc. ISBN 978 – 1- 449- 38259-9

Website References

16s Ribosomal RNA and Universal Primers – Duke University, Durham
URL: http://www.lutzonilab.net/primers/page604.shtml (2011.07.25)

454 Life Sciences - Roche Comapany, Connecticut
URL: http://www.454.com/ (2011.07.25)

Ribosomal Database Project (RDP)– Michigan State University, Michigan
URL: http://rdp.cme.msu.edu/ (2011.07.25)

Quantitative Insight Into Microbial Ecology (QIIME) – Knight Lab, Colorado State University, Colarado
URL: http://qiime.sourceforge.net/ (2011.07.25)

RNA2D Map – University of Texas, Austin
URL: http://www.rna.ccbb.utexas.edu/SAE/2A/RNA2DMap/index.php (2011.07.25)

APPENDICES

Appendix I

Activated sludge sample utilized for the present study:

Sample Group	SampleID	Source	Sample Type
GroupA	A1	Domestic Wastewater Treatment Plant	Activated Sludge
	A2	Wastewater Treatment Plant	Activated Sludge
	A3	Wastewater Treatment Plant	Activated Sludge
	A4	Wastewater Treatment Plant	Digested Sludge
	A5	Laboratory Reactor	Activated Sludge
	A6	Laboratory Reactor	Treated Water
	A7	Laboratory Reactor	Activated Sludge
	A8	Industrial wastewater Treatment	Purified DNA
GroupB	B1	Domestic Wastewater Treatment Plant	Activated Sludge
	B2	Domestic Wastewater Treatment Plant	Activated Sludge
	B3	Domestic Wastewater Treatment Plant	Activated Sludge
	B4	Domestic Wastewater Treatment Plant	Activated Sludge
	B5	Domestic Wastewater Treatment Plant	Raw Water
	B6	Laboratory Reactor	Treated Water
	B7	Laboratory Reactor	Activated Sludge
	B8	Industrial wastewater Treatment	Purified DNA
GroupC	C1	Domestic Wastewater Treatment Plant	Activated Sludge
	C2	Domestic Wastewater Treatment Plant	Activated Sludge
	C3	Domestic Wastewater Treatment Plant	Activated Sludge
	C4	Domestic Wastewater Treatment Plant	Digested Sludge
	C5	Laboratory Reactor	Activated Sludge
	C6	Laboratory Reactor	Treated Water
	C7	Laboratory Reactor	Activated Sludge
	C8	Industrial wastewater Treatment	Purified DNA
GroupD	D1	Domestic Wastewater Treatment Plant	Activated Sludge
	D2	Domestic Wastewater Treatment Plant	Activated Sludge
	D3	Domestic Wastewater Treatment Plant	Activated Sludge
	D4	Domestic Wastewater Treatment Plant	Activated Sludge
	D5	Laboratory Reactor	Activated Sludge
	D6	Domestic Wastewater Treatment Plant	Purified DNA
	D7	Laboratory Reactor	Activated Sludge
	D8	Industrial wastewater Treatment	Purified DNA
GroupE	E1	Domestic Wastewater Treatment Plant	Activated Sludge
	E2	Domestic Wastewater Treatment Plant	Activated Sludge
	E3	Domestic Wastewater Treatment Plant	Activated Sludge
	E4	Domestic Wastewater Treatment Plant	Activated Sludge
	E5	Laboratory Reactor	Treated Water
	E6	Domestic Wastewater Treatment Plant	Purified DNA
	E7	Domestic Wastewater Treatment Plant	Activated Sludge
	E8	Domestic Wastewater Treatment Plant	Activated Sludge
GroupF	F1	Domestic Wastewater Treatment Plant	Activated Sludge
	F2	Domestic Wastewater Treatment Plant	Activated Sludge
	F3	Domestic Wastewater Treatment Plant	Activated Sludge
	F4	Domestic Wastewater Treatment Plant	Treated Water
	F5	Laboratory Reactor	Activated Sludge
	F6	Laboratory Reactor	Activated Sludge
	F7	Laboratory Reactor	Activated Sludge
	F8	Laboratory Reactor	Activated Sludge
GroupG	G1	Domestic Wastewater Treatment Plant	Treated Water
	G2	Domestic Wastewater Treatment Plant	Treated Water
	G3	Domestic Wastewater Treatment Plant	Treated Water
	G4	Domestic Wastewater Treatment Plant	Treated Water
	G5	Laboratory Reactor	Treated Water
	G6	Laboratory Reactor	Treated Water
	G7	Domestic Wastewater Treatment Plant	Treated Water
	G8	Domestic Wastewater Treatment Plant	Treated Water
GroupX	Ax	Domestic Wastewater Treatment Plant	Activated Sludge
	Bx	Domestic Wastewater Treatment Plant	Treated Water
	Cx	Domestic Wastewater Treatment Plant	Treated Water
	Dx	Domestic Wastewater Treatment Plant	Activated Sludge
	Ex	Industrial wastewater Treatment	Purified DNA
	Fx	Industrial wastewater Treatment	Purified DNA
	Gx	Industrial wastewater Treatment	Purified DNA
	Hx	Industrial wastewater Treatment	Purified DNA

Appendix II

The details sequence data obtained from RDP quality analysis process.

Sample: Group of samples
Barcode: attached barcode
Total: Total number of sequences generated at 454 pyrosequencer for each barcode
Selected: selected sequences with barcode attached
Avg. Size: average size of fragment

Sample	Bar Code	Total	Selected	Avg. Size
A	AAAA	17728	959	429
	AATT	24109	1142	429
	ATAT	20206	1011	435
	ATTA	11971	550	440
	TTTT	2823	166	409
	TTAA	4424	281	378
	TATA	3670	180	437
	TAAT	3321	131	426
	No tag	33535	1288	447
	Total	121787	5708	
B	AAAA	14143	1788	438
	AATT	20610	2276	445
	ATAT	0	0	0
	ATTA	13848	1777	425
	TTTT	2614	210	462
	TTAA	1950	247	386
	TATA	3770	341	442
	TAAT	3221	330	438
	No tag	30097	1921	455
	Total	90253	8890	
C	AAAA	13408	1638	439
	AATT	16205	1640	443
	ATAT	20163	1629	435
	ATTA	1762	497	381
	TTTT	3358	188	378
	TTAA	1148	92	434
	TATA	3448	405	433
	TAAT	3563	224	418
	No tag	11961	412	420
	Total	75016	6725	
D	AAAA	11281	1566	436
	AATT	13932	1659	437
	ATAT	17991	1984	439
	ATTA	6634	940	430
	TTTT	2655	191	392
	TTAA	7117	894	447
	TATA	3452	424	438
	TAAT	4505	559	453
	No tag	21522	2172	458
	Total	89089	10389	

Sample	Bar Code	Total	Selected	Avg. Size
E	AAAA	17518	1460	444
	AATT	18472	1811	445
	ATAT	20103	1697	442
	ATTA	14048	1266	435
	TTTT	747	74	398
	TTAA	6706	528	437
	TATA	4347	445	445
	TAAT	6341	462	452
	No tag	26866	1881	461
	Total	115148	9624	
F	AAAA	15422	1779	436
	AATT	14356	1464	432
	ATAT	18675	1951	428
	ATTA	5102	310	419
	TTTT	1618	105	416
	TTAA	4517	324	395
	TATA	3637	348	430
	TAAT	4522	418	439
	No tag	23203	1572	443
	Total	91052	8271	
G	AAAA	21009	1503	431
	AATT	3840	290	408
	ATAT	15396	694	401
	ATTA	10450	933	423
	TTTT	5433	281	440
	TTAA	1932	122	463
	TATA	2440	146	420
	TAAT	5155	306	457
	No tag	16453	799	435
	Total	82108	5074	

Appendix III

Activated sludge samples were collected from four Sequencing Batch Reactors (SBR) operated during the time of experiment.

Reactor A:
Sludge mainly composed of excess sludge collected from previous reactor operation and feeding with synthetic wastewater.

Reactor cycle: Anaerobic (~ 1hour), aerobic (~ 2 hours) and settling (~ 1hour)
Sludge volume: 10L
6 cycles per day

Reactor B:
Sludge mainly composed of excess sludge collected from previous reactor operation and feeding with synthetic wastewater.

Reactor cycle: Anaerobic (~ 1hour), aerobic (~ 2 hours) and settling (~ 1hour)
Sludge volume: 10L
6 cycles per day

Reactor C:
Sludge taken from conventional activated sludge treatment process and feed with mineral and acetate solution (EBPR reactor)

Reactor cycle: Anaerobic (~ 1hour), aerobic (~ 2 hours) and settling (~ 1hour)
Sludge volume: 10L
6 cycles per day

Reactor D:
Sludge taken from conventional activated sludge treatment process and operational conditions ere set up for granular sludge formation

Reactor cycle: Anaerobic (~ 1hour), aerobic (~ 2 hours) and settling (short settling time)
Sludge volume: 10L
8 cycles per day