Analysing the effect of industrial and urban polluted zones on microbial diversity in the SaiGon -DongNai river system (Vietnam)

Thi Tuyet Nga Nguyen

université PARIS-SACLAY

NNT: 2017SACLS582

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Analyzing the effect of industrial and urban polluted zones on microbial diversity of the SaiGon-DongNai river system (Vietnam)

Thèse de doctorat de l'Université Paris-Saclay
préparée à l' Université Paris-Sud

École doctorale N°577: Structure et Dynamique des Systèmes Vivants
Spécialité de doctorat: Sciences de la Vie et de la Santé

Thèse présentée et soutenue à Orsay, le 20 Décembre 2017, par

Mme THI TUYET NGA, NGUYEN

Composition du Jury:

Mme Jacqui, Shykoff Directrice de Recherche, CNRS	Président
M. Dominique, Schneider Professeur, Université Joseph Fourier Grenoble	Rapporteur
Mme Marie-Claire, Lett Professeure, Université de Strasbourg	Rapporteur
M. Jean-Luc, Jung Maître de Conférences, Université de Bretagne Occidentale et Université Bretagne Loire	Examinateur
M. Michael, DuBow Professeur, Université Paris-Sud	Directeur de thèse

Titre : Analyse de l'effet des activités humaines, agricoles, industrielles et domestiques sur la diversité microbienne du système fluvial Saigon-Dong Nai (Vietnam).

Mots clés : diversité microbienne, sédiment, rivière, pollution

Résumé : Le système fluvial Saigon-Dong Nai (SG-DN) est la plus importante source d'eau pour les douze villes et provinces du sud du Vietnam. Il est aujourd'hui gravement pollué par les activités humaines, agricoles, industrielles et domestiques, constituant une menace pour la vie de millions de personnes. Le ministère vietnamien des Ressources naturelles et de l'environnement a rapporté que les rivières ont reçu environ 1,54 milliard de litres d'eaux usées provenant de 70 parcs industriels par jour, dont 35% de déchets médicaux non traités, et que des tests effectués depuis 2006 ont montré des niveaux élevés de pollution, en particulier de substances toxiques organiques. Jusqu'à présent, il n'y a pas de données sur la diversité microbienne dans le système fluvial SG-DN, en particulier dans les sédiments, où la plus grande partie de la biomasse microbienne est généralement localisée.

Les échantillons de sédiments ont été recueillis, réseau hydrographique national SG-DN, à 13 endroits dans les rivières représentant des emplacements pollués. Afin de caractériser les populations microbiennes présentes sur nos sites choisis, l'ADN total des échantillons environnementaux a été extrait et amplifié dans les régions V3 à V1 de l'ADNr 16S. L'étude a révélé que la population microbienne changeait de l'amont vers l'aval au niveau du phylum, du genre et de l'OTU après avoir traversé la zone de population industrielle et dense. De plus, les canaux du bassin versant SG-DN sont fortement pollués par de fortes concentrations de composés organiques (PAH) et possèdent différentes communautés bactériennes par rapport aux échantillons des rivières.

Université Paris-Saclay

Titre : Analyzing the effect of industrial and urban polluted zones on microbial diversity of the SaiGon - DongNai river system (Vietnam).

Mots clés : microbial diversity, sediment, river, pollution

Résumé : The SaiGon-DongNai (SG-DN) river system is the most important major water source for all twelve Southern Vietnam cities and provinces and is now dramatically polluted by industrial and living activities, giving “a threat” to the lives of millions people sharing this water source. The Ministry of Natural Resources and Environment of Vietnam reported that the rivers received around 1.54 billion liters of waste water from 70 industrial parks per day, including 35 percent of untreated medical waste, and tests since 2006 have found pollution in this river has increased to “serious levels”, an especially high concentration of organic toxic substances. Until now, there is no data on the microbial diversity in SG-DN river system especially in the sediments, where most of the microbial biomass is generally located.

The sediment samples were collected in 13 locations across the rivers representing warning polluted locations done by Mr. Nguyen Thanh Hung of the National Water Qualifying in SG-DN river system. In order to characterize the microbial populations present at our chosen sites, the total DNA from the environmental samples were extracted and amplified at the V3 to V1 regions of the 16S rDNA.

The study revealed that microbial population changed from upstream to downstream at the phylum, genus and OTUs levels after running through the industrial and dense population zone. Moreover, the canals of the SG-DN river catchment are heavily polluted with high concentrations of organic compounds (PAHs) and possessed different bacterial communities compared to the samples from the rivers.

Espace Technologique / Immeuble Discovery
Route de l’Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

ACKNOWLEDGEMENTS

I would like to thank my thesis committee members; Dr. Jean-Luc Pernodet, Leader of the Microbiology Department (I2BC), without his presence, my thesis would not be finished; Prof. Pierre Capy, Dean of the Doctoral School SDSV, for the communication between the Doctoral School and its PhD student; Dr. Yves Dessaux, Deputy Director of Microbiology Department (I2BC), for spending time both in the meeting and reading my manuscript, your critical advice helped me a lot to improve the manuscript; Dr. Jacques Oberto, Leader of Cellular Biology of Archaea Laboratory, for the value suggestion.
I also wish to say thank to my jury members; Dr. Jacqui Shykoff, Prof. Marie-Claire Lett, Dr. Jean-Luc Jung and Prof. Dominique Schneider for attending my thesis defense. With all of your value questions, we had a great discussion.
My thanks to Prof. Le Phi Nga and Dr. Nguyen Ngoc Vinh from Vietnam National University, Ho Chi Minh City, for your enthusiasm of the SaiGon-DongNai river project.
I wish to say thank my thesis director, Prof. Michael DuBow. The time I spent with you was hard, but I have learned how to be a good scientist. Thank you, Mike.

Thank you, my beloved colleagues, Jorge Osman, Wang Yang, Gustavo Ribero, Sandra Concha Guerrero and Christophe Regeard, for the wonderful time we spent together.
My thanks to my friends in IGM, Adeline PK, Jerzy Witwinowski and Evelyne Marguet for spending time to attend my thesis defense. Matteo Cossu and Cecile Ounette for your great wishes. My thanks to Roland and Josephine Rebois, for hosting me in their house to finish my manuscript.

I would like to say thanks to my family, especially my mom and dad, who continuously support me. My special thank to *Mi Macho*, who always believes that I can make it to the end of my Ph.D. My dear friends; Libera Latino, for the crazy time we spent together; Aaron Millan Oropeza and Diana Couto for always support me in my Ph.D. My *angel* Muriel Decraene, who helped beyond her ability, my *brother* Aristide Irie and for a special one in my heart, to my dear friends in Viet Nam; A13, A week-Holiday and my students, with their care for my career as a Ph.D student.

My thanks to Campus France, which gave me the opportunity to do my Ph.D in France.

That's here. That's home. That's us. On it everyone you love, everyone you know, everyone you ever heard of, every human being who ever was, lived out their lives… Like it or not, for the moment the Earth is where we make our stand.

Carl Sagan, 1934-1996

For me, our Home is so beautiful and we need to protect it with all of our capacities.

Table of contents

List of Figures:

List of Tables:

List of abbreviation:

APS :	adenosine 5’ phosphosulfate
ARDRA :	amplified rDNA restriction analysis
BOD :	biochemical oxygen demand
CCD :	charge couple device
COD :	chemical oxygen demand
dATPaS :	deoxyadenosine alfa-thio triphosphate
DDE :	1,1-Dichloro-2,2-bis(p-chloro-phenyl) ethylene
DDT :	dichlorodiphenyl-trichloroethane
DGGE :	denaturant-gradient gel electrophoresis
DONRE :	Environmental Management Division of Department of Natural Resources & Environment
EDTA :	ethylenediaminetetraacetic acid
EE2 :	ethinyl estradiol
emPCR :	emulsion PCR
USEPA :	United States Environmental Protection Agency
EPZ :	export-processing zones
GC/MS :	gas chromatography/mass spectrometry
HCMC :	Ho Chi Minh City
IPs :	industrial parks
ISQG:	Interim Sediment Quality Guidelines
LOD :	limit of detection
ND :	non detectable
NDMA :	N-nitrosodimethylamine
OTUs :	Operational Taxonomic Units
PCA :	principal component analyses
PCR :	polymerase chain reaction
PE:	paired-end
PEL :	probable effects level
PAHs :	polyaromatic hydrocarbons
PCBs :	polychlorinated biphenyls
PCD :	Pollution Control Department
PTP :	PicoTiterPlate device

PPi :	pyrophosphate
SAWACO :	Saigon Water Supply Company
SG-DN :	Sai Gon-Dong Nai
STPP :	sodium tripolyphosphate
RFLP :	restriction fragment length polymorphism
rrn operons :	ribosomal RNA
TEL :	threshold effects level
TSS :	total suspended solids
TN :	total nitrogen
TOC :	total organic carbon
USGS :	United States Geological Survey
VNU-HCM :	Vietnam National University of Ho Chi Minh City
WWTP :	wastewater treatment plant

CHAPTER 1: GENERAL INTRODUCTION

1.1. Introduction to the rivers:

Rivers are one part of the Earth's water cycle (also called the hydrologic cycle). Water on Earth exists in 4 main forms: saline water (mostly in the ocean), fresh water (in lakes, rivers, and underground), ice, and atmosphere water (**Fig. 1.1**).

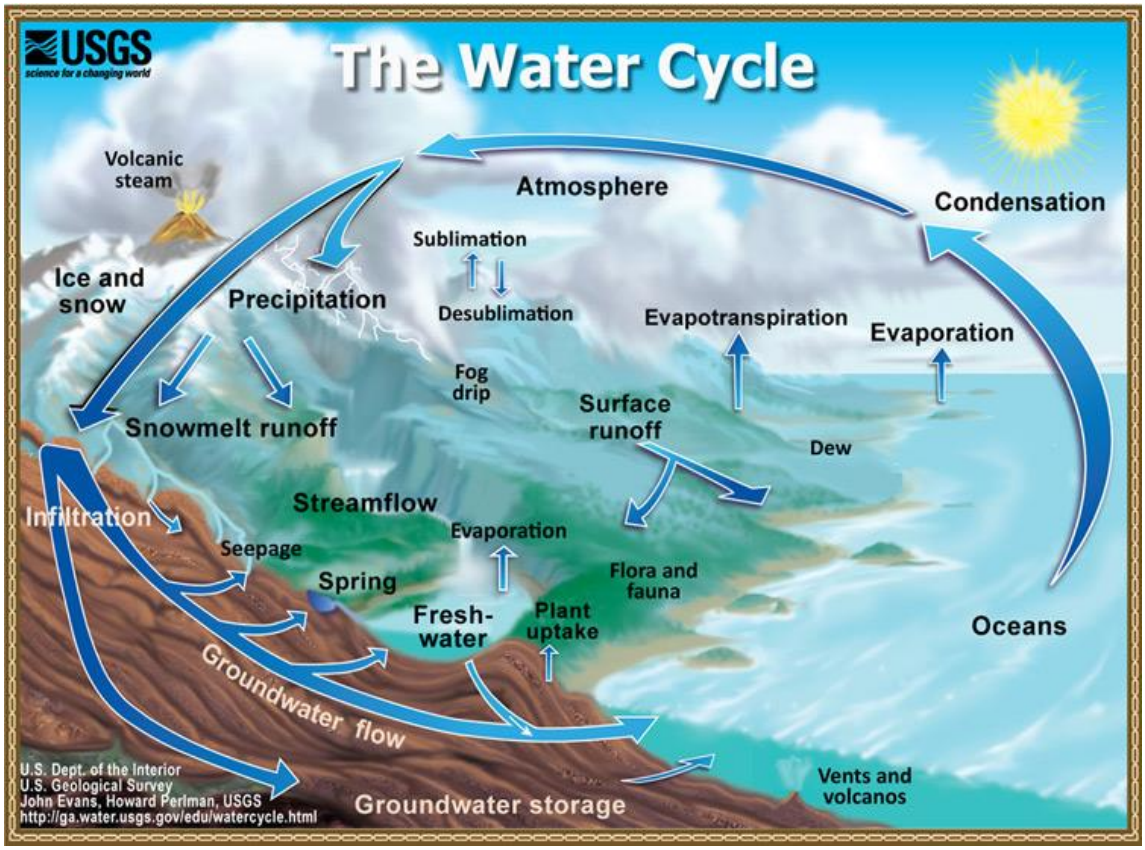


Figure 1.1: Water cycle on Earth **(1)**.

Among these forms, saline water is the most abundant, as ¾ the Earth's surface is covered by oceans. The definition of the liquid water form, such as saline or fresh water, is based on the concentration of NaCl that is present. According to the United States Geological Survey (USGS), there are four categories: fresh water, slightly saline water, moderately saline water and highly saline water **(2)**.

Rivers are a continuous flow of liquid water from one location such as a mountain, towards another location such as a sea, lake, or even another river. The source of a surface river comes from the precipitation of evaporate water (rain), solid water (ice), and other sources of liquid water (e.g. ground water) **(3)**.

1.2. Introduction to water pollution:

1.2.1. Definition:

Water pollution is the contamination of natural water bodies by chemicals, physical parameters, radioactive and even by microorganisms **(39)**. Contamination can be found in water reservoirs such as lakes, rivers, oceans, aquifers and groundwater. Water pollution occurs when substances called pollutants are introduced into water bodies and cause negative effects on surrounding wildlife and habitats. Water pollution affects the entire biosphere, including plants, animals and other organisms living in these bodies of water **(39)**.

There are two kinds of pollution sources: one is natural phenomena such as volcanoes, algal blooms, storms, and earthquakes, while the other is from human activities (also called anthropogenic*)* resulting from industrial wastewater discharges, chemical substances of agricutural lands and urban waste water. These have different impacts on physical, chemical and biological features in water. Physical changes can comprise elevated temperature and discoloration. The alteration of physical and chemistry in water can include acidity (changes in pH), electrical conductivity, temperature, and eutrophication **(40)**.

1.2.2. Water pollution issue of the world:

Water pollution is reported in urban and industrial areas all around the world, including China **(16, 17)**, India **(18)**, the USA (**19, 20)**, Mexico, Brazil, Chile **(21, 22, 23)**, Morocco **(24, 25)**, France plus Spain **(26, 27, 28)**, and Thailand plus Laos **(29, 30, 31)**. In 2005, according to Qiu Baoxing, Chinese Deputy Minister of Construction, 90 % of the water in China's cities and 75 % of its lakes suffer from some degree of pollution **(32)**.

Polluted water has been one of the main causes of various health problems in humans throughout the world. In the year 2000, the estimated mortality due to water sanitation hygiene-associated diarrheas and other water sanitation associated diseases, such as schistosomiasis, trachoma, intestinal helminth infections, was 2,213,000 people **(33)**. About 65 million people suffer from fluorosis, a crippling disease due to high amounts of fluoride, while 5 million are estimated to be due to arsenicosis in West Bengal due to high amounts of arsenic **(34).**

In China, in 2007, nearly 500 million people lacked access to safe drinking water **(35)**. More than 130 residents of two villages in the Guangxi Province in southern China were poisoned by arsenic-contaminated water **(16)**. In addition to health problems in

humans, contaminated water also causes adverse effects on aquatic species, such as ethinyl estradiol (EE2) directly and indirectly affects phytoplankton and zooplankton **(36, 37)**.

Water shortages are also the consequence of water pollution due to increasing human populations. Water shortages have been a constant worry for China for centuries. In 2005, according to Qiu Baoxing, more than 100 of China's 660 cities face 'extreme water shortages'. China supports 21 % of the world's population with just 7 % of its water supplies **(32)**. In China, about 75 % of the population (approximately 1.1 billion people) are without access to unpolluted drinking water, according to China's own standards **(33)**. Of the 632 districts examined to determine the quality of ground water, only 59 districts had water safe enough to drink. Bagalkot town in the Karnataka state of India has the most unsafely drinking water, with 5 out of the 6 pollutants exceeding safety limits **(38)**.

1.2.3. Pollutants:

1.2.3.1. Definitions:

'Such a substance has to be present in the environment beyond a set or tolerance limit, which could be either a desirable or acceptable limit. Hence, environmental pollution is the presence of a pollutant in the environment; air, water and soil, which may be poisonous or toxic and will cause harm to living things in the polluted environment' (41). Alternatively, 'A pollutant is any substance in the environment, which causes objectionable effects, impairing the welfare of the environment, reducing the quality of life and may eventually cause death' **(42).**

1.2.3.2. Types of pollutant:

1.2.3.2.1. Organic pollutants:

Some pollutants are biodegradable and therefore will not persist in the environment. Biodegradation is the chemical dissolution of materials, either by biological means or abiotic decomposition of organic material by microorganism **(43)**. Biodegradable matter is generally organic material that serves as a nutrient for microorganisms. Some industrial pollutants, such as hydrocarbons (e.g. oil), polychlorinated biphenyls (PCBs), polyaromatic hydrocarbons (PAHs) are substances that can be biodegraded **(44)**. However, the degradation products of some pollutants are themselves polluting. Insecticides such as dichlorodiphenyltrichloroethane (DDT) can be biodegraded to produce 1,1-dichloro-2,2-bis(p-chlorophenyl) ethylene (DDE) **(340, 341, 342, 343)**. According to the United States Environmental Protection Agency (EPA), DDE is a toxic pollutant that can cause tumors in mice, hamsters and rats. DDE has been classified as a Group B2 probable human carcinogen.

Other organic water pollutants detergents (e.g. sodium tripolyphosphate (STPP), ethylenediaminetetraacetic acid –EDTA), disinfection by-products (DBPs) in disinfection of drinking water such as N-nitrosodimethylamine (NDMA), which is a possible human carcinogen. Food processing waste from industry includes organic substances such as proteins, carbohydrates, lipids, suspended solids, biochemical oxygen demand (BOD) and chemical oxygen demand (COD) substances **(45)**. BOD and COD substances are decomposed material such as food waste (anthropogenic), dead plant or animal tissue which are substances for micro-organisms during the decomposition process leading to the reduction of oxygen in aquatic environments **(46)**.

1.2.3.2.2. Nutrient pollutants:

Nutrient pollution means pollution caused by nutrients that can support the growth of terrestrial plants close to the water body, or weeds and algae in the water. Fertilizers for agriculture, wastewater and sewage contain high levels of nutrients. Fertilizers provide additional N, P, K and other elements in the form of inorganic chemical compounds such as ammonium nitrate (NH_4NO_3), urea ($CO(NH_2)_2$), calcium ammonium nitrate ($Ca(NO_3)_2 \bullet NH_4NO_3 \bullet 10H_2O$), phosphorus ($P_2O_5$), calcium dihydrogen phosphate monohydrate ($Ca(H_2PO_4)_2$), monoammonium phosphate ($NH_4H_2PO_4$), diammonium phosphate ($(NH_4)_2HPO_4$, K_2O), water-soluble salts of metals such as copper sulfate ($CuSO_4$), or in the form of organic compounds such as decayed plant matter and animal waste. Wastewater or sewage containing human waste also brings organic nutrients to water, providing additional nutrients for autotrophic organisms such as plants, algae, weeds and cyanobacteria. These organisms can then grow rapidly and consume the oxygen in the water, causing oxygen depletion and death for other obligate aerobic organisms such as fish **(47, 48)**. Moreover, the overabundance of plants and algae can cover the water surface and prevent sunlight, thus encouraging the growth of anaerobic bacteria and altering the normal conditions of the ecosystem in the water body. This process is also called eutrophication. The profuse growth of plants decreases water clarity, leading some species to form unsightly scums, while certain species of algae cause taste and odor problems in drinking water. Some blue-green algae can be toxic to animals. Another consequence of eutrophication is altering the species composition of a river ecosystem, with native biota being displaced in the environment. Finally, changes in nutrient content can also indirectly affect river chemistry, such as the amounts of carbon dioxide being uptaken and released by plants can alter the pH in the water **(49)**.

1.2.3.2.3. Non-toxic pollutants:

Particular concentrations of chemical substances such as calcium, sodium, iron and manganese occur naturally in aquatic systems. However, the concentration of these elements can be increased due to natural phenomena or human activities, leading to adverse affects on aquatic environments **(40)**.

1.2.3.2.4. Microbial pollutants:

Microbial contaminants caused by pathogenic bacteria, viruses, protozoa and helminths are the most common and widespread health risk associated with drinking water (WHO, 2004. Guidelines For Drinking Water Quality 3rd Ed. page 123).

Disease-causing microorganisms are referred to as pathogens. Although the vast majority of bacteria are either harmless or beneficial, a few pathogenic bacteria can cause disease to humans, plants or animals. For example, gram-negative bacteria, such as *Burkholderia pseudomallei,* found in soil and water, can cause the infectious disease in human and other animals called melioidosis **(50)**. This organism is often found in tropical countries such as India, Thailand and northern Australia **(51, 53, 53)**. Infections by *B. pseudomallei* can develop into kidney disease, blood disease, heart disease and other fatal disorders **(54, 55)**. High levels of pathogens may result from on-site sanitation systems (septic tanks, pit latrines) or inadequately treated sewage discharge. This can be caused by a sewage plant designed with fewer secondary treatment facilities (more typical in less-developed countries) **(33)**.

Other waterborne microbial pathogens, besides bacteria, include virus and protozoa. Some common waterborne pathogens include bacteria such as *Salmonella typhi* **(56)**, *Escherichia coli* **(57)**, *Vibrio cholera* **(58)**, *Pseudomonas aeruginosa* **(59)**, *Shigella spp.* **(60)**, parasitic *Cryptosporidium* **(61)**, *Giardia lamblia* **(61)** and *Norwalk virus* **(62)**.

1.2.3.2.5. Heavy metal pollution:

"Heavy metals" is a general collective term which applies to the group of metals and metalloids with atomic density greater than 4 g/cm^3, or 5 times or more, and greater than water (42). Heavy metals are elements such as lead (Pb), cadmium (Cd), zinc (Zn), mercury (Hg), arsenic (As), silver (Ag) chromium (Cr), copper (Cu), iron (Fe), boron (Bo) and platinum (Pt) **(41, 42, 63)**. The sources of these elements can be both natural and anthropogenic, such as mining or industrial activities. In nature, heavy metals occur as natural constituents of the earth's crust. In addition, as elements, they cannot be degraded or destroyed and therefore persist in the environment as their ores in different chemical forms, from which they are recovered as minerals **(41, 42, 63)**.

The major causes of metal pollution is from anthropogenic sources such as mining operations **(42, 64)** and they are emitted both in elemental and compound (organic and inorganic) forms **(42)**. The various industrial point sources include former and present mining sites, foundries and smelters, combustion by-products and emission from vehicles **(42, 65)**. Heavy metals are widely used in agriculture, industry and for medical purposes **(42)**. An excess of heavy metals from these activities can cause different levels of effects on ecosystems **(63)**. In water, they leach into underground waters, moving along water pathways and eventually depositing in the aquifer, or are washed away by run-off into surface waters, thereby increasing the pollution levels in water and soils **(42)**.

Some heavy metals are trace elements and have biological importance such as Fe and Zn. However, concentrations of these trace elements beyond acceptable limits can have a severe impact on human or ecosystem organisms **(42)**. The toxicity of heavy metals can cause organ damage, cancer and neurological damage to humans, depending on the dose **(66, 67)**. Heavy metals will combine with biomolecules such as proteins and enzymes and form stable biocompounds, altering their structures and thus preventing them from normal function **(67)**.

1.3. Tools for studying environmental bacteria:

1.3.1. 16S rDNA:

The 16S rDNA gene is a 1550 base pair DNA sequence that is present in all bacterial genomes. The transcript of this gene, called 16S rRNA, has a structural and functional role as a part of the ribosomal small subunit. The ribosome is the translation machine that occurs in all living organisms. It is composed of a large (50S) and a small (30S) subunit. The 16S rRNA binds to 21 proteins to form the small subunit (30S) of the ribosome (**139**). Due to its vital function in bacteria, the 16S rRNA gene is highly conserved. For this reason, in 1987 Woese and his colleagues chose the 16S rRNA gene to measure the evolutionary relationships among bacteria **(140, 141)**.

1.3.2. 16S rDNA and bacterial identification:

The 16S rDNA gene sequence is used in at least two major applications: (i) identification and classification of isolated pure cultures and (ii) estimation of bacterial diversity in environmental samples without culturing through metagenomic approaches **(142)**.

The 16S rDNA gene has nine variable regions as 'informative regions' for measuring evolutionary relationships among bacteria **(143)**. Along with these nine variable regions are the conserved regions **(Fig. 1.2)**.

Before next-generation sequencing (NGS) was invented, studying the 16S rDNA gene was often based on prior cloning. The 16S rDNA gene was sequenced and assigned to the appropriate taxonomic groups. However, the numbers of sequences generated by cloning are not sufficient for estimating the totality of bacterial community in environmental samples based on the number of reads. In contrast, NGS is capable of generating from thousands to millions of sequences, ideal for studying environmental bacterial diversity. However, the sequence length generated from NGS is limited from 100-500 nt, depending on the sequencing platform. Sequencing the 16S rDNA gene using NGS techniques challenges microbiologists with two questions (i) what is the appropriate region of 16S rDNA gene to be studied? and (ii) what are appropriate universal primers to use? This leads to the question of whether using the short sequences of 16S rDNA are enough to accurately assign taxonomy. To answer the first question, many studies have examined the taxonomic assignment accuracy among each hypervariable region of the 16S rDNA gene or a combination of two or three hypervariable regions. Studies showed that no single region is sufficient to accurately distinguish among bacteria **(145, 146)**, due to different levels of variation across the nine hypervariable regions of the 16S rDNA gene **(147)**. The combination of two or three variable regions of 16S rDNA is generally found to be sufficient for identifying bacteria at most taxonomic levels **(145)**. For example, hypervariable regions V1-V3 or V3-V5 are usually used to study 16S rDNA in the human microbiome **(148)**. In pathogenic bacteria diagnostic assays, choosing which hypervariable region to analyze is critical. Region V1 was shown to the best to distinguish between *Staphylococcus aureus* strains, as is region V2 among *Mycobacterial* species and V3 among *Haemophilus* species. Regions V2 and V3 were also found to be suitable to differentiate bacteria to the genus level, while region V6 can also be used for members of the *Enterobacteria*. These hypervariable regions are found to be more informative than V4, V5, V7 and V8 in pathogenic bacteria classification **(145)**. However, the full-length sequence still provides the most accurate identifications in 16S rDNA based methods **(144, 145, 149).**

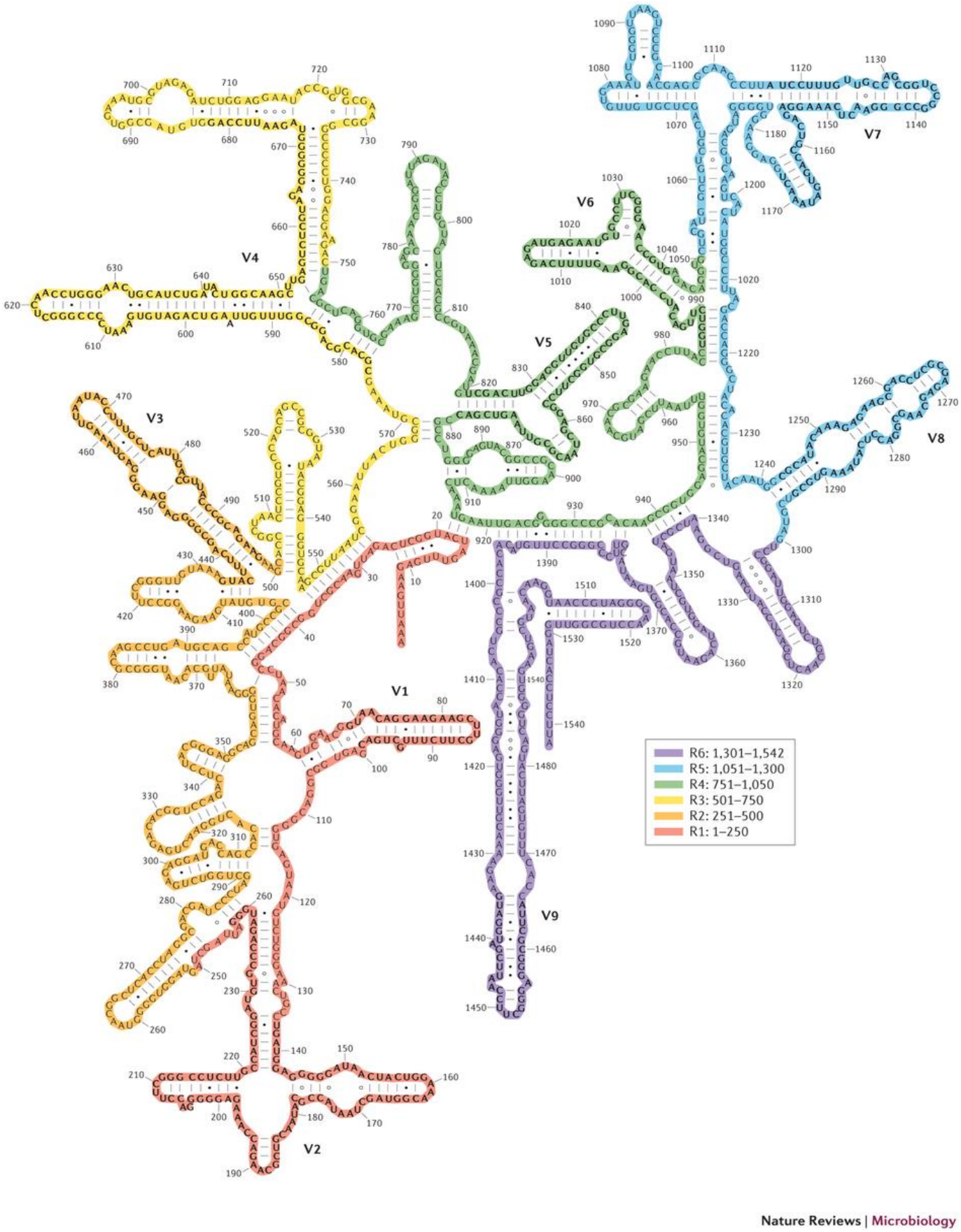


Figure 1.2. Secondary structure of 16S rRNA with nine hypervariable regions V1-V9 (in bold letters) **(144).**

The second question is what are appropriate universal primers to use? As previously discussed, the common cloning universal primer spans the whole length of the 16S rDNA gene. For example, the primer pair 27F and 1492R (based on *E. coli* positions) was designed by Weisburg et al. in 1991 and were evaluated by Frank et al. 2008 **(150, 151).** With the use of next-generation sequencing, sequencing 16S rDNA gene can thus target particular hypervariable regions. For examples, Titanium 454 pyrosequencing, producing 500nt sequences, led to the design of primer pairs 27F and 518R for the V1-V3 region and suitable primer pairs for the V3-V4 region (**Fig. 1.3**) **(152, 153).**

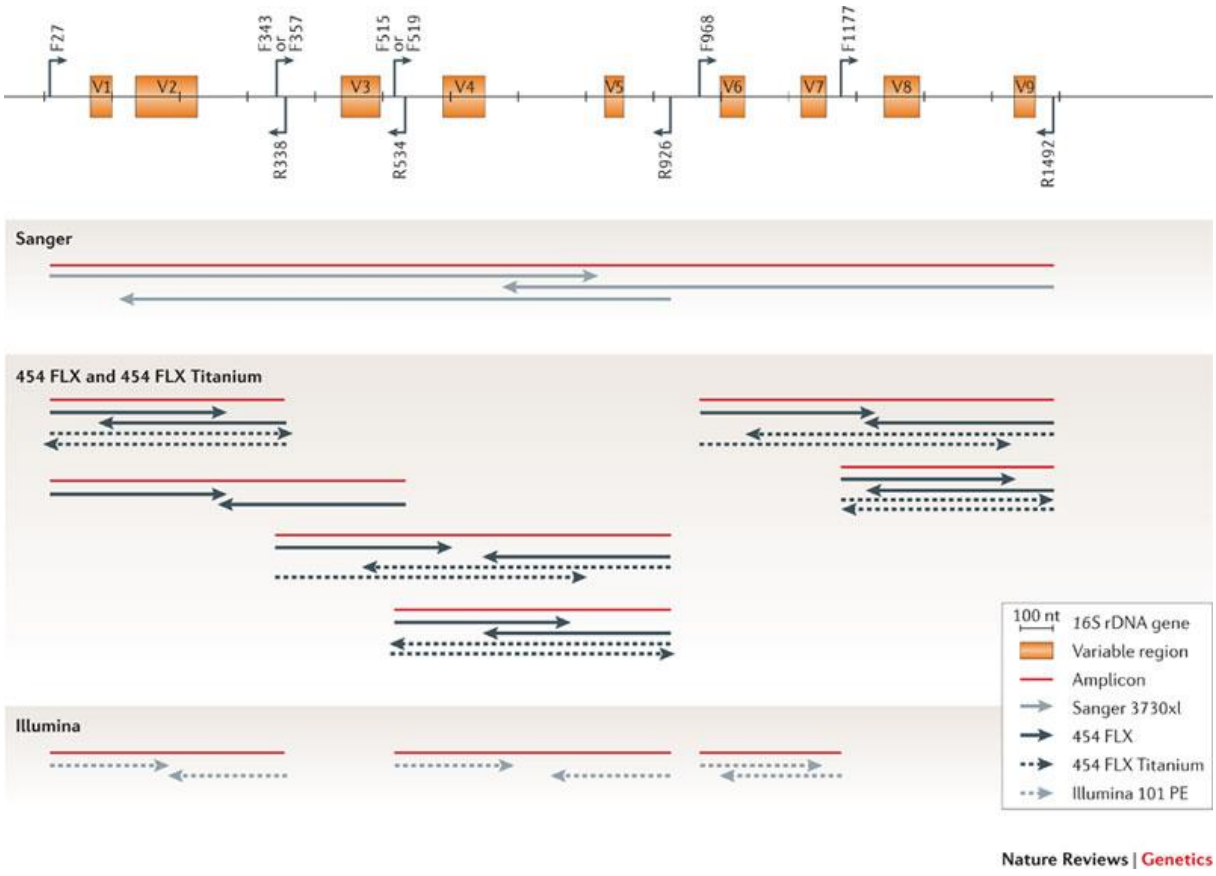


Figure 1.3. Taxonomic profiling consists of generating a PCR amplicon (in red) of the (partial) 16S ribosomal RNA (rRNA) gene (top) with selected PCR primers, followed by sequencing that amplicon with a preferred technology (grey arrows) : Sanger ABI 3730xl, 454 (FLX and FLX Titanium) and Illumina 101 paired-end (PE) sequencing technologies **(152).**

A number of studies have been published to find the 'best' universal primer pair for amplifying 16S rDNA and fully access total bacterial diversity in environmental

samples. There are several criteria established for selecting 16S rDNA primers in amplicon sequencing **(153)**:

1/ Primer pairs should be located in the highly conserved regions of the 16S rRNA gene.

2/ The hypervariable region of the amplicon is suitable to identify bacteria community members in the given samples.

3/ The amplicon should be compatible with the sequencing strategy depending on their read length (cloning versus next-generation sequencing, single or paired-end).

4/ The forward and reverse primers can be modified for amplification reactions.

First, primer pairs ought to be located in the conserved regions of the 16S rRNA gene so that they are able to amplify a wide range of bacterial community members. Second, as discussed above, different hypervariable regions are suitable for different bacterial communities. For this reason, the effect of primer choice on bacterial taxonomic identification has been studied. Bacteria in mammals such as steer rumen in the human gut can be accurately classified at the genus level with most primer pairs, while a grassland soil microbial community appears compatible with primer pairs 341F and 1406R. This may be due to the dependence of variable coverage of the 16S rDNA genes in reference databases **(154)**. Thirdly, there are two NGS techniques commonly used in studying 16S rDNA: 454 pyrosequencing and Illumina. Current studies have shown that the V3-V4 and V4-V5 regions yield the highest classification accuracy for both 454 and Illumina technologies. However, 454 pyrosequencing appears more appropriate for studying 16S rDNA because it provides longer read length and less error than Illumina **(155)**. Mori et al. (2014) found that the non-degenerate primer pairs 342F-806R, spanning regions V3-V4, may be the best for studying prokaryotic 16S rRNA genes without amplifying some eukaryotic community members. Many studies proved this primer pair spanning region V3-V4, is the best choice for studying bacterial communities in different environments **(156, 157, 158, 159).**

1.3.3. 16S rRNA and Metagenomics:

In the 1980's, sequencing of 16rRNA genes directly from environmental samples was begun in order to analyze natural microbial populations, particularly by the groups of Stahl **(160)**, Lane **(161)** and Pace **(162)**. Since then, microorganisms in natural environments, such as the picoplankton communities in marine ecosystems **(163)**, sulfate-reducing bacteria in sandy marine sediments **(164)**, bacterial communities in Siberian tundra soils **(165)**, and uncultivated hot spring cyanobacterial, chloroflexus-like and spirochete-like mat inhabitants **(166)**.

This has opened a whole new era in microbial ecology. In 1995, Rudolf Amann and colleagues estimated that >99% of microorganisms observable in nature typically cannot be cultivated using standard techniques **(167).**

The genomes of hundreds of organisms from all three domains of life (archaea, bacteria and eukarya), as well as those of viruses, have now been sequenced from the environment of interest using random shotgun sequencing approaches **(168)**. Other conserved marker genes, informative and suitable for phylogenetic analyses, have been studied, including the bacterial large subunit ribosomal gene (23S RNA), genes involved in translation (elongation factor EF-Tu), genes encoding the β subunit of bacterial RNA polymerase (rpoB) and genes participating in DNA repair processes (recA) **(169, 170)**. Shotgun sequence data and marker genes together answer two main questions of environmental microbiologists, which are 'who is there?' and 'what might they be doing?' **(171)**. According to Handelsman, metagenomic (also referred to as environmental and community genomics) is the genomic analysis of microorganisms by direct extraction and analysis of DNA from an assemblage of microorganisms **(172).**

Bacteria identification through sequencing of marker genes, such as the 16S rDNA gene, or through whole-genome sequencing, can answer the question 'who is there?'. To answer the question 'what might they been doing', functional (metabolic) studies of microbial communities are required. One can look at the homology among genes of interest with the existing genes in databases, such as genes for the degradation of plant matter or operons concerned with potassium metabolism **(173)**. In shotgun sequencing, DNA from the environment is randomly sequenced and assembled, if possible, through overlapping regions. As larger sequences are built, functional genes of interest can be used to construct the genomes of uncultured microorganisms. Metabolic pathway information, such as special nutritional requirements or biogeochemical functions, carbon and nitrogen metabolism, energy acquisition, and how the microbial communities interact with each other and within the environment can be potentially extracted from this information **(174, 175)**. Furthermore, to fully access the functional microbes in different environments, metatranscriptomics and metaproteomics have been developed, opening a new era of ecology by looking at microbes to fully understand how an ecosystem can work **(176)**.

1.3.4. Techniques for studying 16S rDNA:

In order to study the diversity and ecology of a particular bacterial community, several methods have been developed. They include culture-dependent and culture-independent analyses **(177)**. The culture-dependent analyses involve the enrichment and

isolation of bacterial community members in the desired environment and the culture-independent analysis involves sequencing and pattern analysis of the amplified 16S rDNA gene from the directly extracted DNA. Among them, pattern analysis approaches allow rapid and sensitive detection of bacterial diversity, and include Amplified rDNA Restriction Analysis (ARDRA) **(178, 179, 180)**, Denaturing Gradient Gel Electrophoresis (DGGE) **(181)**, Temperature-Gradient Gel Electrophoresis (TGGE) **(182)**, Restriction Fragment Length Polymorphism (RFLP) **(183)**.

1.3.4.1. Pattern analysis:

The purpose of pattern analysis is to evaluate banding patterns of the amplified 16S rDNA gene products on gels **(184, 185).**

1.3.4.1.1. Denaturing gradient gel electrophoresis (DGGE) & Temperature gradient gel electrophoresis (TGGE):

Denaturing gradient gel electrophoresis (DGGE) is a molecular fingerprinting method that separates PCR-generated DNA products according to the differences in their sequence G-C content **(184)**. DNA fragments from a sample containing multiple organisms are amplified using PCR **(186)**. The PCR products are then subjected to increasingly higher concentrations of chemical denaturant in a polyacrylamide gel in constant temperature (about 60ºC) during the process **(185, 186)**. The higher G/C content of double stranded-DNA (dsDNA) molecule requires a greater denaturant due to their higher hydrogen bond energy, therefore, resulting in differential mobility of DNA molecules on the gradient gel **(184)**. Differing G/C content sequences of DNA from different bacteria will denature at different denaturant concentrations resulting in a pattern of bands, with each band theoretically representing a different bacterial population present in the community **(187)**. The brightest bands in a DGGE profile are often assumed to represent the dominant members of the community **(185)**. TGGE techniques are similar to DGGE with a temperature gradient being used rather than a denaturing chemicals gradient.

a) Advantages:

One of the advantages of DGGE is its sensitivity for detecting even single base-pair differences between amplicons **(185)**. For that reason, DGGE can be used to distinguish between mutated and wild-type sequences without prior knowledge of what these sequences are **(186)**.

b) Disadvantages:

A DGGE/TGGE experiment covers < 400 bp fragments of 16S genes and requires large quantities of DNA for effective resolution **(185)**. Due to the existence of multiple

copies of rRNA genes in a single organism, multiple bands for a single species can occur and lead to ambiguity of the results in DGGE **(185)**. In addition, this method can be difficult to apply to extremely complex communities that produce hundreds of bands on a DGGE profile, which become difficult to visualize individually **(185)**. PCR-DGGE fingerprinting has been widely used in environmental microbiology to detect the similarities and differences of the dominant populations of microbial communities and was the most commonly used method of community characterization in the literature, appearing in roughly 15% of articles surveyed in 2005 **(185).**

1.3.4.1.2. Restriction fragment length polymorphism (RFLP) & Amplified ribosomal DNA restriction analysis (ARDRA):

Another pattern analysis technique is Restriction Fragment Length Polymorphism (RFLP), or Terminal Restriction Fragment Length Polymorphism (T- RFLP). These methods are well developed for rDNA genes such as 16S or 23S and can be called Amplified Ribosomal DNA Restriction Analysis (ARDRA) **(185)**. The rDNA genes are first amplified by PCR from the total extracted DNA. The amplified community DNA's are then digested by various restriction enzymes. The operating principle of RFLP is that divergences in the rDNA gene sequences of different species will create differences in restriction sites for various enzymes, therefore, creating different patterns for each species on gel electrophoresis. If the correct restriction enzymes are used, what can emerge is a unique fingerprint for each species or strain **(185)**. This digested DNA is run on a gel, producing a pattern of fragment sizes that is characteristic of the community. For single isolates or clones, the digests can be run on regular agarose. However, in studies of complex communities, the large number of DNA fragments produced by this method can often only be resolved using polyacrylamide gels **(185).**

a) Advantages:

One of the microbiology advantages of ARDRA is that it is rapid and cost-effective so that most molecular labs can perform this method. ARDRA can be performed directly on PCR-amplified community DNA or on clones from a library of PCR-amplified DNA **(185, 188)**. T-RFLP/ ARDRA and DGGE/TGGE can both be recommended for pattern analysis without further sequencing if a non-heterogeneous gene is used, and with additional sequencing using heterogeneous genes **(184)**.

b) Disadvantages:

A technical limitation of ARDRA is that each experiment requires optimization before performing **(185)**. However, not all restriction enzymes will serve equally well for

the analysis. Another problem that is that if a restriction site occurs mainly within highly conserved regions of the ribosomal genes, then many of the fragments produced from different species can be difficult to distinguish from one another **(185)**. For that reason, choosing the appropriate regions of ribosomal genes for ARDRA analysis should be examined.

1.3.4.2. Sequencing:

There are at least four major options when selecting a sequencing platform for metagenomic studies including dideoxy sequencing (Sanger), pyrosequencing (454 – Roche), SOLiD™(Applied Biosystems), and Illumina® (formerly known as Solexa) **(189)**.

The Sanger sequencing method was developed by Frederick Sanger, who shared the Nobel Prize in Chemistry in 1980 and is considered as 'first-generation' sequencing **(190)**. Sanger sequencing can produce sequences up to 750-1000 nt. The newer methods, including Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences are referred to as next-generation sequencing (NGS) **(190)**. One of the advantages of next-generation sequencing is the ability to produce an enormous volume of data cheaply — in some cases in excess of one billion short reads per instrument run **(190)**.

1.3.4.3. Pyrosequencing:

1.3.4.3.1. Introduction:

Pyrosequencing is a DNA sequencing technique that utilizes enzyme-coupled reactions and bioluminescence to monitor the pyrophosphate release accompanying nucleotide incorporation, in real-time **(191)**. It is based on the "sequencing by synthesis" principle and different from Sanger sequencing which is chain termination with dideoxynucleotides developed by Frederick Sanger and colleagues in 1977 **(192, 193)**. Pyrosequencing was developed by Mostafa and Pal Nyren at the Royal Institute of Technology in Stockholm in 1996 **(194, 195, 196)**.

1.3.4.3.2. 454 Life Science:

454 Life Sciences was founded by Jonathan Rothberg in June 2000 in Branford, Connecticut (USA), originally as 454 Corporation, a subsidiary of Roche Applied Science company, specializing in high-throughput DNA sequencing. 454 Life Sciences was awarded the Wall Street Journal's Gold Medal for Innovation in the Biotech-Medical category in 2005 **(196)**. In the same year, 454 Roche released the GS20 sequencing machine – the first next-generation DNA sequencer on the market. In 2008, 454 Sequencing developed the Genome Sequencer FLX instrument along with GS FLX

Titanium series reagents, with the capacity to sequence 400-600 million nt per run with 400-500 nt read lengths **(190, 191)**.

1.3.4.3.3. Mechanism of 454 pyrosequencing:

The 454 sequencing principle involves three steps: DNA library preparation, emulsion (em PCR) and the running step (**Fig. 1.4**).

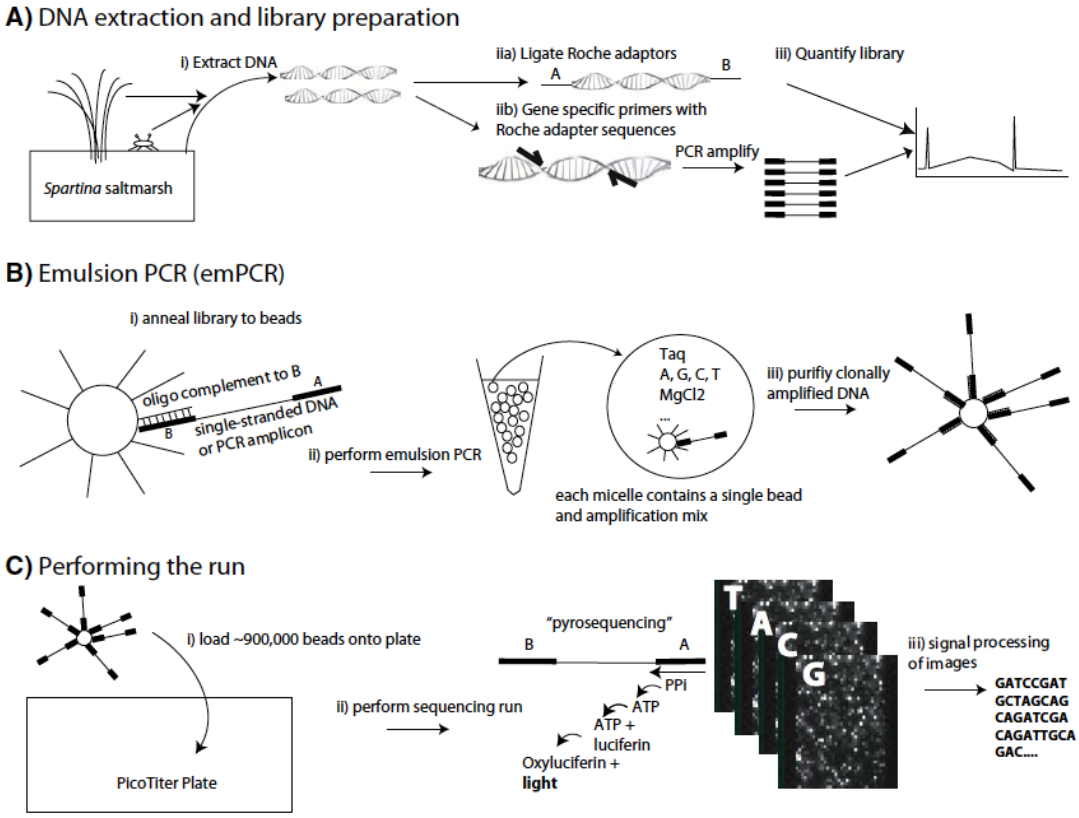


Figure 1.4. The schematic portrayal of the Roche/454 Life Science workflow **(197)**.

Step 1. DNA library preparation:

Total DNA is extracted from environmental samples (soil, sediment, feces, water, drinking water network biofilms, etc.). Then, 16Sr DNA genes are amplified with fusion primers to create an amplicon DNA library with appropriate DNA fragment length between 400-600 bp **(197)**. Fusion primers are oligonucleotides that comprise adaptor sequences, barcode sequences (provided by 454 Roche company) and 16S rDNA universal primers (**Fig. 1.5**).

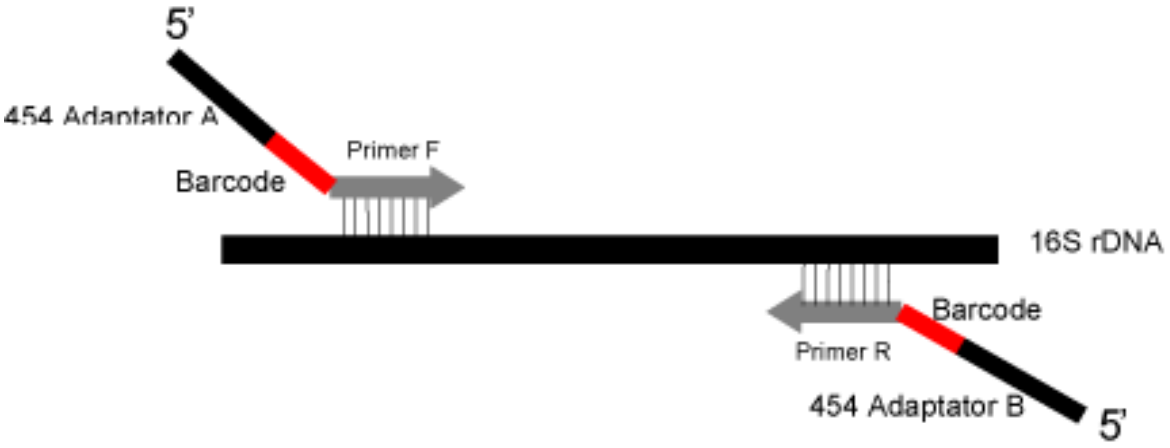


Figure 1.5. Standard fusion primers designed by 454 Roche company **(153)**.

Step 2. emPCR:

Emulsion PCR (emPCR) is a PCR reaction in which a single-stranded DNA template (from DNA library preparation), thermostable DNA polymerase and PCR reagents, and million of oligos complement to any adaptor B are attached to a bead as a primer; together located in an oil:water micelle droplet. Typically, most droplets will contain only one single-stranded DNA (sst) template. During the PCR reaction, the adaptor B from ssDNA will bind its complement oligos on the bead. Then, DNA polymerase synthesizes the complementary strand of the DNA template. At the end of the PCR reaction, the bead contains approximately ten million identical copies of ssDNA template fragments. Those beads that do not contain DNA (called null beads) will be eliminated by enrichment steps. There are about approximately 2.4 million beads in a 454 sequencing reaction **(197, 198, 199)**.

Step 3. Running step:

i) Loading step:

The sequencing process takes place in a microfabricated glass plate called a PicoTiterPlate device (PTP) which contains 1.6 million picoliter reactor wells. Four bead layers are loaded into PTP sequentially including layer 1: enzyme beads pre-layer, layer 2: amplified DNA from emPCR and packing beads, layer 3: the enzyme beads post-layer and layer 4: PPiase beads. The loaded PTP device is centrifuged at 4000 RPM for 10 min at room temperature for each loaded bead layer, and the beads are deposited into the wells according to the Sequencing Method Manual (GS Junior Titanium Series, May 2010). The diameters of the wells are designed so that only a single capture bead (sstDNA bead) will fit into each well **(198)**.

ii) Reaction step:

After loading with all beads, the PTP device is placed into the sequencing machine. The beads containing four different enzymes necessary for the pyrosequencing process, including DNA polymerase, ATP sulfurylase, luciferase and apyrase. Then, a solution of an individual deoxyribonucleotide triphosphate (dTTP, dATP, dGTP or dCTP) is added into the wells sequentially in a fixed order (T,A,G,C). In the pyrosequencing process, each time a dNTP solution is added is called a flow and an order of four flows of T, A, G, C is called a cycle. In total, there are 800 flows for each dNTP, divided into 400 cycles in the 454 GS Junior and 454 FLX systems. If one of the added dNTP's is complementary to the template DNA strand in the bead, DNA polymerase incorporates that deoxyribonucleotide triphosphate into the template strand and releases a pyrophosphate (PPi). Then, ATP sulfurylase converts PPi and Adenosine 5' phosphosulfate (APS), already present in the wells, into ATP. Finally, luciferase enzyme uses the produced ATP to convert the substrate luciferin to oxyluciferin. Oxyluciferin generates visible light that is detected by a charge coupled device (CCD) camera. If the dNTP is not complementary to the template strand, apyrase degrades it into adenosine monophosphate (AMP) and inorganic phosphate leading to no light produced and no signal on the CCD camera. The next dNTP is then added. It should be noted that in the DNA polymerase reaction, the natural deoxyadenosine triphosphate (dATP) is substituted by deoxyadenosine alfa-thio triphosphate (dATPaS) as a substrate of DNA polymerase but not the substrate of luciferase. This substitution ensures that luciferase does not produce a false light signal in the dATP flow of the sequencing process **(198, 199, 200, 201)**. The height of each light signal peak is proportional to the number of nucleotides incorporated.

iii) Data analysis:

This process goes into 3 main steps (**Fig. 1.6**, **1.7**):

- Image capture
- Imaging processing
- Signal processing. In this step, there are 2 ways to filter the data: **Shotgun filter** versus **Amplicon filter.**

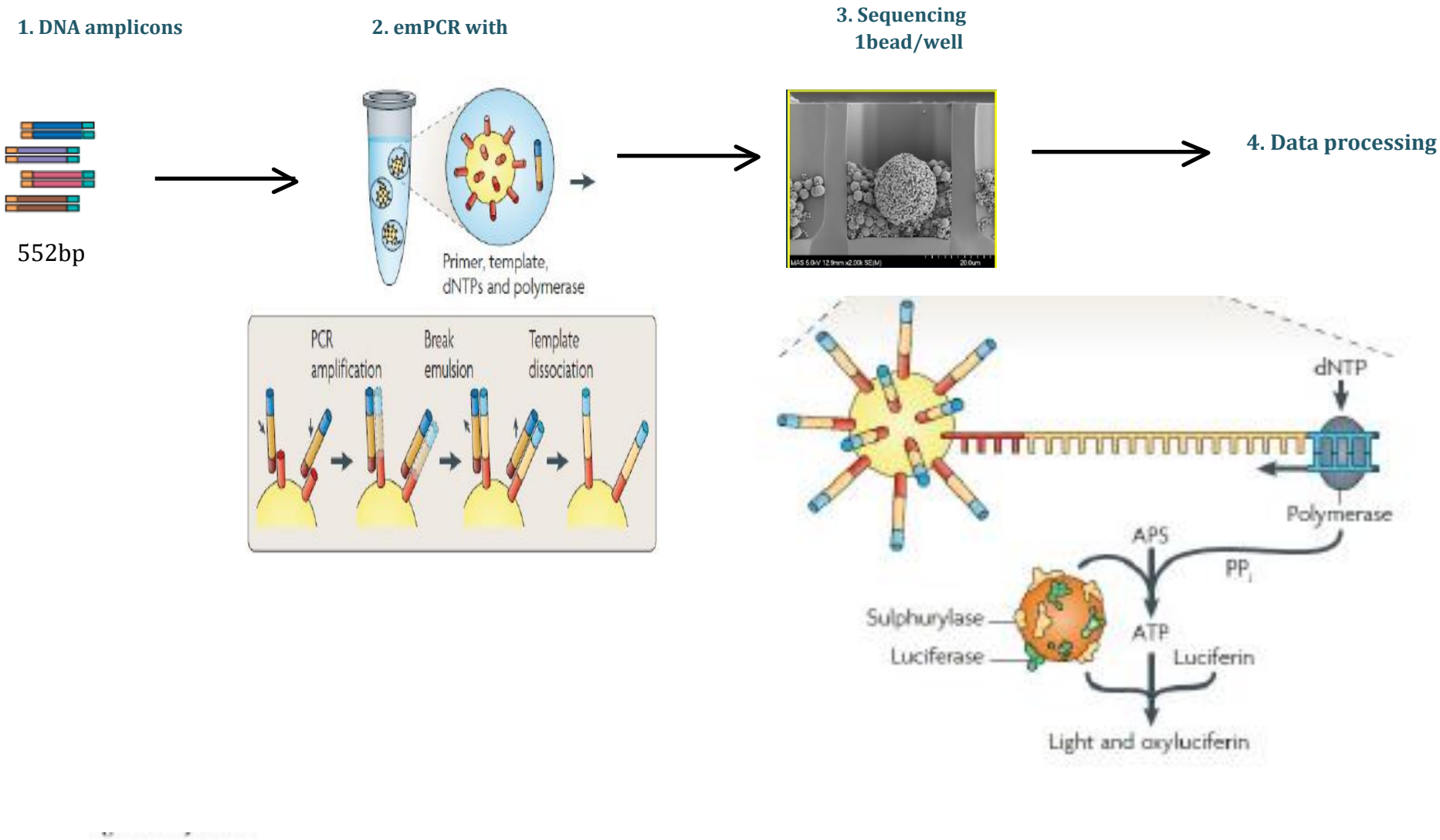


Figure 1.6. 454 GS Junior/FLX Titanium sequencing processing **(260)**.

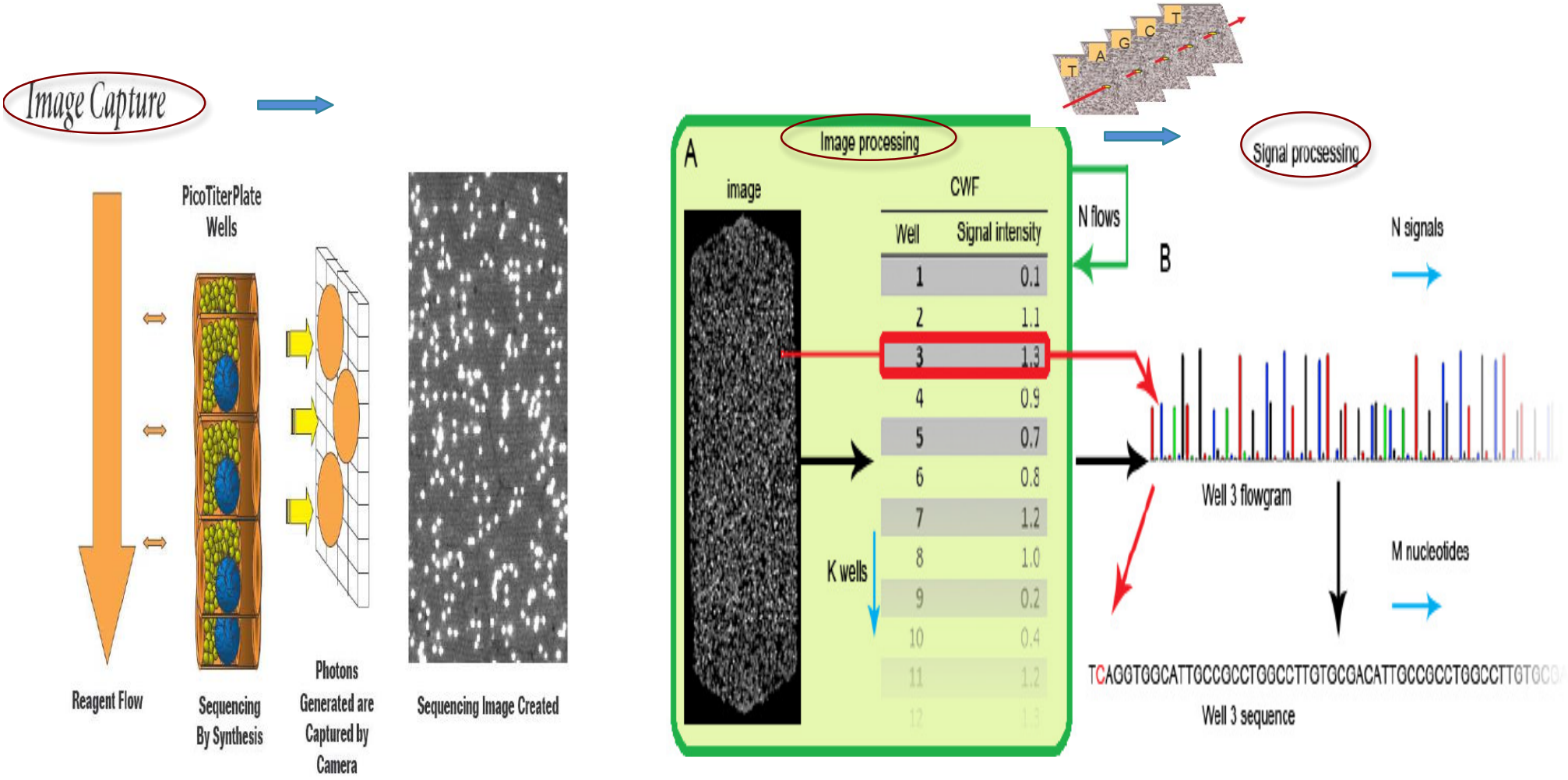


Figure 1.7. The data processing steps include 3 main steps: Image Capture, Image Processing and Signal Processing (**261**).

1.4. Limits of bacterial community analysis based on 16SrDNA approaches:

Several disadvantages of bacterial community analysis based on 16S rDNA approach have been reported including 1) sample collection 2) cell lysis procedures 3) PCR amplification 4) 16S rDNA copy numbers and 5) DNA sequencing errors. These have been proved to affect the estimation of microbial composition in the environmental studies.

1.4.1. Sample collection:

Bias caused by sample collection affects the composition of the microbial communities. Xiong and colleagues (2012) found bacterial distribution change among geographic distance along with pH in sediments of Tibetan Plateau **(202)**. Bacterial communities can also change in depth **(203, 204).** A study of Bacteria and Archaea in distances ranging from 0.01m to 1000m suggest that geographic sampling replication should be taken into account when studying microbial diversity in environments to avoid spatial variation **(204).**

1.4.2. Cell lysis procedures:

The cell lysis and DNA extraction processes can be problematic, as humic acids present in soil and sediment samples often coextract with nucleic acids and can inhibit downstream DNA processing reactions **(205)**. Contamination with humic acids can decrease soil community estimations **(206, 207)**. Many attempts had been used to remove humic substances, such as bead beating extraction, polyvinylpolypyrrolidone (PVPP)-sepharose 2B column elution **(208)** or adding mannitol in the lysis buffer **(209)** and using commercial methods such as Norgen's Soil DNA Isolation Kit **(210).**

1.4.3. PCR amplification:

Several factors of PCR reactions such as the cycle numbers, DNA polymerase, DNA template features from total environmental DNA extraction, as well as primer-pairs designed from 16S rDNA gene conserved regions, have been shown to be able to distort the proportion of indigenous bacterial communities in the studied samples **(211, 212, 213).** Increasing PCR cycles, template concentrations and polymerase errors introduce noise to data analysis, leading to the overestimation of the true diversity **(218, 219)**. PCR amplification of many different homologous genes, such as multigene families a single species or genes coding for rRNA, can generate sequence artifacts including chimeras and heteroduplexes **(212, 213).**

Heteroduplexes are formed by the cross-hybridization from two different double-stranded DNA molecules **(214)**. The phenomenon is generated when primer:template ratios

decrease towards the end of PCR reactions. At that point, low concentrations of primer lead to inefficiency of primer-annealing to DNA template. The concentration of PCR products increases, resulting two single-stranded DNA molecules from 2 different double-stranded DNA hybridizing molecules to form heteroduplexes. The concentration of heteroduplexes can also increase as template diversity is increased **(214).**

Chimeric molecules are formed when an incompletely extended PCR product acts as a primer on a heterologous sequence **(215)**. In PCR reactions, the polymerases in some circumstances do not completely synthesize a DNA template molecule (e.g. stem-loop block) **(215)**. The highly homologous truncated molecule can hybridize to other DNA template molecules, forming a sequence that has a part of a DNA template from one species and the other part of DNA template from another species (**Fig.1.8**) **(217)**. Heteroduplexes and chimera artifacts are the pitfalls in the comparative studies of genes that have high homology levels such as 16S, 18S, 23S ribosomal DNA and other marker genes such as EF-Tu (**215, 216)**.

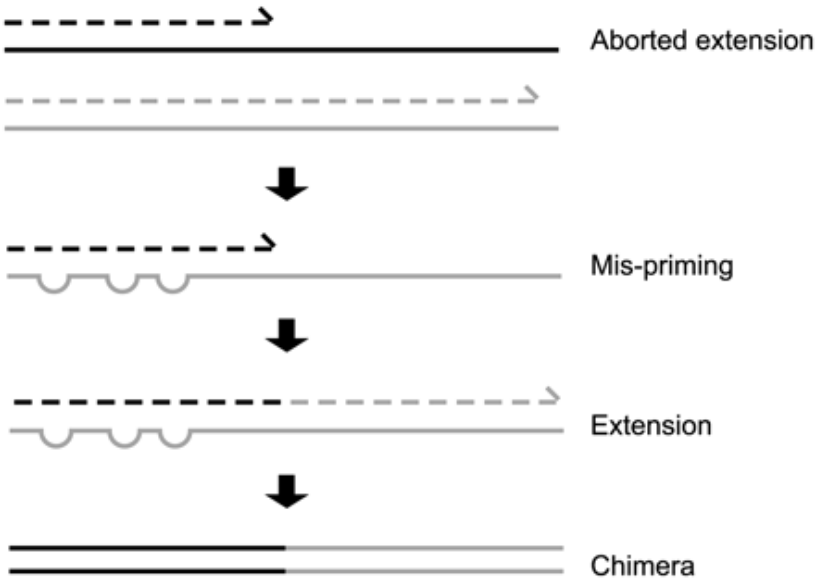


Figure 1.8. Chimera sequences formation **(217)**.

Increasing PCR cycles often cause chimera formation, leading to the overestimation of bacterial richness **(218)**. Liesach et al. (2003) showed that chimera formation could result in the overestimation of bacterial community diversity **(219)**. Moreover, errors caused by different thermostable DNA polymerases can lead to errors such as indels (insertions and deletions) or mismatches (misincorporation of bases) **(220, 221)**. To eliminate the bias caused by heteroduplexes in mixed-template PCR products, Thompson et al. (2001) developed a process called recondition PCR. The mixed-template

PCR products were diluted 10-fold and re-amplified in a low number of cycles (3 times) **(214)**.

Qiu et al. (2001) minimized artifacts generated by PCR, such as the presence of chimeras, mutations and heteroduplexes, by reducing the number cycles **(220)**. Using high fidelity thermostable DNA polymerases can minimize the microbial community analysis bias caused by these issues **(218, 221, 222)**.

1.4.4. Numbers of 16S rDNA copies:

The ribosomal RNA operon (rrn) coding for 16S, 23S and 5S rRNAs can be present as single or multiple copies in the genomes of microorganism (bacteria and archaea) **(223)**. The rRNA gene plays an essential role in protein production so that the presence of more than one copy of the rRNA generally correlates with the growth rate of bacteria **(224)**. *E. coli* has 7 copies of the rRNA genes and decreasing the number of rRNA genes by deletion of rrn operons leads to a decrease in growth rate **(225)**. For example, in some fast-growing soil bacteria, multiple copies (average of 5.5 copies) of rRNA operons (rrn) per genome were found **(224)**. In contrast, bacteria that are found in low nutrient areas (generally oligotrophe) contain low copy number (single copy) of the rRNA operon **(226)**. In *Borrelia turicatae*, there is one 16S rRNA gene per cell, with 15 genes per cell in *Brevibacillus brevis*, while in *Clostridium paradoxum* 5 copies are present **(227, 228)**. *Erythrobacter litoralis* also has a single 16S copy **(229)**. Thus, bacteria abundance evaluation should also take into consideration with the relative numbers of 16S gene per cell as bacteria, which contain relatively high 16S gene copy numbers, will appear to be in high abundance, while bacteria with relatively low 16S gene copy numbers will appear to be in low abundance **(227)**.

1.4.5. Pyrosequencing errors:

454 pyrosequencing was originally designed with shotgun genome sequencing in mind. In genome sequencing, multiple reads are assembled to create contigs from consensus sequences. The error rate of each nucleotide position produced which is by this technique is covered by the large number of overlapping reads, also called ‘coverage’. This makes the genome sequencing from high throughput platforms very reliable. The quality filtering processes of pyrosequencing, therefore, do not require very stringent measurement (**199**). Inadequate numbers of reads provided by cloning techniques lead to an underestimation of the true environmental microbial diversity. Amplicon sequencing was developed for 16S rDNA and other marker genes. Consequently, the errors generated by pyrosequencing do not greatly affect genome shotgun sequencing but can have a large

impact on amplicon sequencing. For example, the “rare biosphere” can be a product of sequencing errors **(230)**.

As described in the mechanism of 454 pyrosequencing, light signals are produced and transferred into ‘letter’ DNA sequences. The intensity of the light increases in proportion to the number of nucleotides incorporated in the sequencing reaction. This process is named base calling **(199)**. The base calling process was reported to have errors from several sources **(231)**. The first cause is the number of identical nucleotides that are incorporated, called homopolymers (a stretch of identical nucleotides, for example, AAAA or GGGGG). Error rates are accumulated as the homopolymer length increases **(199, 231)**. The 454 platforms are also found to have errors such as nucleotide insertions and deletions (indels) more frequently than substitutions (**Fig. 1.9**) **(231, 232)**. Moreover, 454 pyrosequencing was found to have high errors toward the 3’ end of a sequence **(232, 233)**.

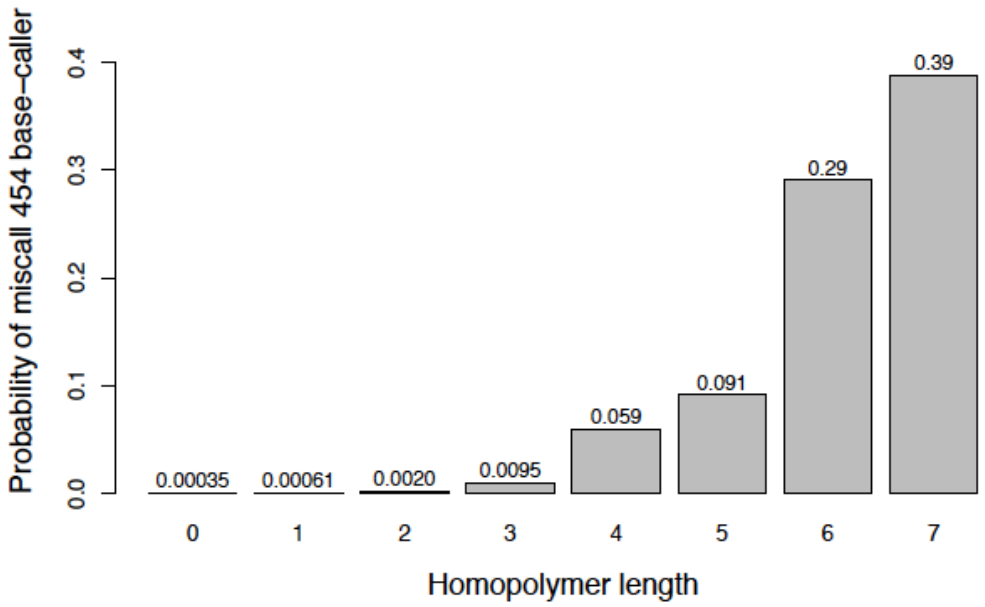


Figure 1.9. Probability of miscalls by the native 454 base-callers on homopolymers (**231**).

An accumulation of 454 sequencing errors at the 3’end of the reads may be due to chemical exhausted reaction. After 800 flows, chemical by-products accumulate in the wells, leading to inefficiency of sequencing reactions. Cross-contamination of the wells (called interdiffusion) can also occur. In the pyrosequencing process, reagents flow across the PTP. The flow dynamics transport the chemical reagents in wells to their neighbors and can increase the background noise by approximately 10% **(199)**. This could explain why errors accumulate according to the spatial location of PTP **(233, 234)**.

Another reason for errors is the high similarity sequence of 16S rDNA amplicons. The 16S rDNA gene consists of the conserved regions, which are highly similar to all bacteria. When conserved regions of 16S rDNA amplicons are sequenced, all the wells in PTP light up at the same time. This can cause cross-contaminated light signals among the wells, giving them more light intensity and leading to insertion errors.

1.5. Bacteria in various environments:

1.5.1. Bacteria in natural environments:

Bacterial community composition has been studied in many different ecosystems on Earth. To answer the question how different environments affect the bacterial composition, microbial ecologists, when studying bacterial diversity, generally first focus on two questions: what bacteria are present (via the particular phyla, classes, orders, families, genera and species) and in what relative proportions? For example:

1.5.1.1. Freshwater sediments:

Studies of surface freshwater sediments of a shallow eutrophic lake by 16S rRNA gene cloning method revealed several bacterial phyla with different proportions **(83)**. In this study, a total of 112 clones from the 16S rRNA gene library were analyzed and revealed 12 phyla including *Proteobacteria*, *Nitrospira*, *Acidobacteria*, *Bacteroidetes*, *Chlorobi*, *Actinobacteria*, *Cyanobacteria*, *Verrucomicrobia*, *Planctomycetes*, *Chloroflexi*, *OP8*, *WS3*. Among those phyla, members of the *Proteobacteria* is the most abundant accounting for 47% of the total number of clones, followed by *Nitrospira (*13.4%), *Acidobacteria (*8.0%), *Chloroflexi (*7.1%), *Bacteroidetes (*6.3%) and *Chlorobi (*4.5%). Members of phyla *Planctomycetes*, *Actinobacteria*, *Cyanobacteria*, *Verrucomicrobia*, *OP8* and *WS3* presented with low abundance of 2.7%, 1.8%, 0.9%, 0.9%, 0.9%, respectively. The unidentified bacteria at the phylum level account for 3.6 % of the total number of clones.

For the *Proteobacteria*, members of the Δ-*Proteobacteria* has been indicated as the representative bacterial lineage in benthic environments, since this group was more frequently recovered from sediments than from water columns, in which *α, β,* and γ-*Proteobacteria*, *Bacteroidetes*, and *Actinobacteria* were observed as the dominant groups **(83, 84, 85, 86)**. This finding may be due to the oxidation-reduction potential gradient between the water and sediment environments **(83)**. In this study, as well as previous reports, frequently detected clones were moderately related to strict anaerobes, such as sulfate reducers (the genera *Desulfococcus*, *Desulfomonile*, and *Desulfonema*) and

syntrophic bacteria (e.g, members of the genus *Syntrophus*), within Δ-*Proteobacteria* **(83, 84, 87, 88).**

1.5.1.2. Coastal sediments:

A study of tidal-flat sediments down to 360 cm of depth revealed the presence of phyla *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Spirochaetes*, and *Chloroflexi* **(89)**. Members of γ-*Proteobacteria* were found almost exclusively in the upper sand-dominated interval, whereas *Firmicutes*, *Bacteroidetes*, and *Chloroflexi* were detected mainly within the deepest layers at 220 cm and below **(89).**

Members of β-*Proteobacteria* were found to be dominant and consistent with 10%-29% proportion in various lacustrine environments. In contrast, members of δ-*Proteobacteria* and *Verrucomicrobia* proportion varied from 1% -29% in different location **(90).**

1.5.1.3. Soil:

Collected data of clone libraries of 16S rRNA genes from different regions, including forest soil (from Canada, Brazil, Austria, Germany, Australia), pasture soil (from Brazil, Switzerland, United Kingdom), arid woodland, cropping rotation, wheat, grassland (United States), arid landscape (Australia), moorland (Germany) revealed what appear to be a general composition of soil bacterial communities **(91)**. Phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Verrucomicrobia* were found to be dominant in soil **(91)**. In addition, members of genera *Arthrobacter, Bacillus, Streptomyces, Agrobacterium, Alcaligenes, Nocardia* and *Pseudomonas*, present with proportions of 3%-40%, 5%-45%, 23%-30%, up to 13%, 1%-8%, 3%-10% and 2%-10%, respectively, have been found **(91, 92).** Another study of anoxic rice paddy soil revealed members of *Bacillus*, *Nitrosospira*, *Fluoribacter*, *Acidobacterium* were likely involved in the process of degrading rice straw **(93).** *Bacillus*-like sequences are also present as the predominant bacteria in Dutch grassland soils **(94)**.

At the phylum level, *Acidobacteria* appears to be one of the most abundant phyla in soil habitats. Up to 50% of the sequences obtained in many 16S rRNA gene clone libraries from soil belong to the phylum *Acidobacteria* **(95)**. This suggests that this phylum plays an ecologically important role in soil ecosystems **(96)**. Moreover, the phylum has nearly as deep phylogenetic branchings as that of the *Proteobacteria*, suggesting the metabolic diversity in metabolic terms **(96)**. In addition, molecular analyses of different soils from temperate climate zones using 16S rRNA gene clone libraries showed that most soil bacteria fall into eight major phylogenetic groups, including those identified using

cultivation-based approaches, such as *Proteobacteria,* its subclasses α-, β-, γ-, δ-, ε-, *Firmicutes, Actinobacteria* and *Bacteroidetes* **(96)**. 16S rRNA gene sequences affiliated with the candidate divisions TM6, TM7 and OP11 have so far been obtained from different soil samples, suggesting that these bacterial groups find their home in soil **(96)**. The phylum *Verrucomicrobia* is also globally widespread and abundant in soil, although relatively few *Verrucomicrobia* have been cultivated. Between 7%-21% of the 16S rRNA gene sequences were affiliated with this phylum, with members of the class "Spartobacterium" (subdivision 2) predominant, as well as members of subdivisions 3 and 4 of the *Verrucomicrobia*. Among the isolated *Verrucomicrobia* were members of species of *Prosthecobacter* (subdivision 1) and *Ultramicroba* (subdivision 4). Besides the *Verrucomicrobia,* several cultured representatives were from the *Planctomycetales* and *Acidobacteria* phyla **(96).**

A 100 bp amplicon sequences generated by pyrosequencing from three agricultural and one forest soil types from North and South America revealed phyla *Proteobacteria and Bacteriodetes* with proportion of 40% and between 15%-25% of the sequences affiliated, respectively. In agricultural soils, a significant proportion of the sequences had high similarity to the ammonia-oxidizing Archaea using 100% similarity, only a few genera were detected in all four soil-types, including member of the *Chitinophaga* (*Bacteroidetes*) and *Acidobacterium* (*Acidobacteria*) **(97)**.

Sequences of 1500-bp fragments of 16S rDNA cloned from sugarcane rhizosphere soil include 29.6% *Proteobacteria*, 23.4% *Acidobacteria*, 12.1% *Bacteroidetes*, 10.2% *Firmicutes* and 5.6% *Actinobacteria* **(98)**. The phylum *Verrucomicrobia*, whose prevalence in N-fertilized soils was approximately 0.7% and increased to 5.2% in the non-fertilized soil, suggested that this group might be an indicator of nitrogen availability in soils. At the genus level, *Bacillus* was the most predominant, accounting for 19.7% of all genera observed. Classically reported nitrogen-fixing and/or plant growth-promoting bacterial genera, such as *Azospirillum, Rhizobium, Mesorhizobium, Bradyrhizobium*, and *Burkholderia*, were also found, although at a lower prevalence **(98)**. Bacteria of the phyla *Proteobacteria* and *Actinobacteria* were the most numerous within the rhizosphere, representing 32.1% (59/184) and 42.9% (79/184) of all isolates, respectively **(99)**. The study of bacteria in Brazilian savanna-like vegetation soil by 454 pyrosequencing 16S rDNA spanning the V5-V9 region allowed the identification of 17 phyla. Among them, *Acidobacteria* were dominant in all areas studied with a relative frequency of 40–47%, closely followed by *Proteobacteria* accounting for 34–40% of the

sequences. Five phyla including *Actinobacteria, Verrucomicrobia, Planctomycetes, Gemmatimonadetes,* and *Bacteriodetes* were considered abundant, with sequence frequencies above 1%, and 10 phyla (*Chlamydiae, Firmicutes, OP10, TM7, Chloroflexi, Cyanobacteria, Nitrospira, Spirochaetes, Thermomicrobia,* and *BRC1*) were considered low abundance, with sequence frequencies below 1%. The most abundant class was *α-Proteobacteria,* corresponding to 52–57% of all *Proteobacteria* sequences. Unclassified bacteria represented the third most abundant group in this study, corresponding to 8–13% of all sequences **(100).**

Bacterial soil communities can also be characterized into two different soil types; one is with intensive cultivation (tomato, beans and corn, etc.) and the other is forest soil (unchanged by man), located in Guaíra, São Paulo State (Brazil). The use of 16S rRNA analysis allowed identification of several bacterial populations in the soil belonging to the following phyla: *Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria* and *Verrucomicrobia* in addition to the others that were not able to be classified, and members of the *Archaea* (**101)**. The bacterial composition of cultivated soil presents with *Actinobacteria* (10%), *Bacteroidetes* (6%), *Firmicutes* (34%), *Proteobacteria* (36%), *Verrucomicrobia* (6%) and unclassified bacteria (8%), while in forest soil these phyla are present in different proportions with *Acidobacteria* (44.3%), *Actinobacteria* (4.5%), *Bacteroidetes* (5.7%), *Firmicutes* (4.5%), *Proteobacteria* (19.3%), *Verrucomicrobia* (19.3%). The proportions of phyla *Acidobacteria*, *Firmicutes*, *Proteobacteria* and *Verrucomicrobia* significantly change between cultivated and forest soil. The cultivated soil was dominated with *Firmicutes* (34%) and *Proteobacteria* (36%) while the forest soil was dominated by members belonging to the *Acidobacteria* (44.3%), *Proteobacteria* (19.3%) and *Verrucomicrobia* (19.3%) phyla **(101).**

A study of bacterial communities by 16S rRNA gene clone library analysis in Western Amazon soils also revealed differences between primary forest, old secondary forest, pasture and crop soils. The percentage of sequences assigned at the phylum level are *Acidobacteria* (38.8%), *Actinobacteria* (6.1%), *Bacteroidetes* (8.3%), *Chloroflexi* (0.3%), *Firmicutes* (2.4%), *Gemmatimonadetes* (0.8%) and *Proteobacteria* (36.2%), including representatives of the classes α-, β-, γ- and δ- of *Proteobacteria*. Of these sequences, 7% could not be classified to phyla. Clear differences were observed in community composition, as shown by the differential distribution of *Proteobacteria, Bacteroidetes*, *Firmicutes*, *Acidobacteria* and *Actinobacteria* **(102).**

1.5.1.4. Air:

The presence of phyla, including γ-*Proteobacteria*, *Actinobacteria*, *Bacteriodetes*, and *Firmicutes* can be observed in aerosol samples **(103, 104)**. This was confirmed by a study of indoor air bacteria with *Proteobacteria* (abundance 41 %), *Actinobacteria* (27%), *Firmicutes* (9%), and *Bacteroidetes* (3%) **(105)**. At the genus level, they include *Propionibacterium* with the proportion 12%, *Diaphorobacter* (10%), *Alicyclobacillus* (6%) *Methylobacterium* (4%), *Sphingomonas* (4%), *Hymenoba-ter* (2%), *Pseudomonas* (2%), and *Roseomonas* (1%) **(105)**. Among these genera, sequences affiliated with *Propionibacterium* were found to be the most abundant in outdoor air, skin, pets, carpet, bathtub tiles, and tap water samples, suggesting their association with human activities. Genera such as *Corynebacterium, Staphylococcus*, *Acinetobacter* and *Kocuria* are microbes that typically colonize the skin of humans and other organisms (**105**). In addition, member of the genus *Corynebacterium* and *Staphylococcus* are also found on kitchen countertops and refrigerator samples **(105)**.

Outdoor-related sources (including outdoor air and doorsteps) were dominated by bacteria belonging to the genera *Propionibacterium, Pseudomonas*, *Staphylococcus*, *Sphingomonas* and *Janthinobacterium*. Water-related sources (bathtub tiles, showerheads, tap water, toilets) were characterized by a relatively high proportion of bacteria belonging to *Propionibacterium, Sphingomonas, Methylobacterium* and *Alicyclobacillus* genera. In addition, sources such as skin and saliva were almost exclusively comprised of *Propionibacterium* sp. **(105)**.

Another study of airborne bacterial communities in Norway, Sweden and Finland by sequencing full-length 16S rDNA clones revealed members of genera *Staphylococcus*, *Acinetobacter*, *Pseudomonas*, *Micrococcus*, *Bacillus*, *Pantoea*, *Curtobacterium*, *Enterococcus* and *Stenotrophomonas* **(103)**. In summary, genera, such as *Staphylococcus* and *Micrococcus* and *Bacillus* are commonly found in air samples from different sources **(103, 105, 106, 107, 108, 109, 110)**. Bacteria can come from different sources such as *Acinetobacter* bacteria which are widely found in nature, mostly in water and soil. However, they have also been isolated from the skin, throat, and various other sites in healthy people **(103, 111)**.

1.5.1.5. Oral microbiota:

Andersson et al. (2008) used 454-pyrosequencing to analyze the V6 hypervariable region of 16S rRNA gene (approximately 280) amplified from samples collected from the throat, stomach, and feces biopsies of six healthy individuals (aged 61–76 years), and three

were *H. pylori* positive by culture. The 56,382 reads, with a mean length of 73 nucleotides, were BLAST-searched (similarity 95% identity) against a reference database of more than 90,000 near-full-length 16S rRNA genes from the Ribosomal Database Project (RDP), and 88% of the reads could be assigned to RDP reference sequences and annotated. The vast majority (99%) of the annotated reads belonged to five bacterial phyla: *Firmicutes*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria* and *Fusobacteria* **(112).**

In this study, *Firmicutes* composed most of the bacterial community abundance in feces with 81.2 ± 11.2% and half of the community abundance in throat with 55.6 ± 13.6% and have a significant abundance in *H. pylori* negative stomachs with 29.6 ± 15.9%. The abundance of *Firmicutes* is reduced dramatically in *H. pylori* positive stomachs with least abundant (1.8 ± 0.6%). *Actinobacteria* is most dominant in *H. pylori* negative stomachs (46.8 ± 18.9%), significantly abundant in throat (14.5 ± 3.9%) and fece (14.6 ± 9.8%) and least abundant in *H. pylori* positive stomach samples (1.1 ± 0.7%). *Bacteroidetes* is mediately represented in throat 20.0 ± 8.6%, less abundant in *H. pylori* negative stomachs (11.1 ± 8.7 %), least abundant in feces (2.5 ± 2.6%) and underrepresented in *H. pylori* positive stomach samples (0.8 ± 0.6%). Interestingly, *Proteobacteria* comprise most of the sequences in *H. pylori* positive stomachs, with 96.2 ± 1.8%, while this phylum presents just a significant amount in *H. pylori* negative stomachs (10.8 ± 3.2%), minor composition in throat (4.7 ± 3.4%) and least abundant in feces (1.7 ± 1.5%). *Fusobacteria* has a small composition in throat (5.1± 3.7 %), least abundant in *H. pylori* negative stomach (1.1 ± 1.1 %), underrepresented in *H. pylori* positive stomachs (0.1 ± 0.01) and absent in feces (0%). *Firmicutes* is the dominant phylum in feces and throat, while *Actinobacteria* is the dominant phylum in *H. pylori* negative stomachs while *Proteobacteria* is the dominant phylum in *H. pylori* positive stomachs. *Fusobacteria* are present with a small proportion in these four samples while *Bacteroidetes* are present with a significant amount in throat and *H. pylori* positive stomach samples **(112, 113).**

Bik et al. (2006) used cloned libraries of 16S rRNA gene (position from 8 to 806 spanning the V1-V4 region) to study bacteria from gastric specimens (corpus of the stomach & antrum in 14 subjects) of 23 adults (22 men and 1 woman, mean age 59 years). Among these 23 subjects, 12 were tested positive for *H. pylori* and 1,833 high-quality sequences were obtained for all 23 subjects after removing chimeric, vector, human, and poor-quality sequences. Dominant phyla were *Bacteroidetes* (10%), *Fusobacteria* (5%), *Firmicutes* (25%), *Actinobacteria* (10%), and *Proteobacteria* 10% for non *H. pylori* and 40% for *H. pylori* **(114).**

The *Bacteroidete* abundance of gastric samples here is similar to *H. pylori* negative stomachs of Andersson et al. (2008) study above with proportions of 10% versus 11.1 ± 8.7%, respectively. Similarly, *Firmicutes* abundance presents with proportion of 25% versus 29.6 ± 15.9%. For the *Actinobacteria*, the composition is different, 10% versus 46.8 ± 18.9%, as are the *Fusobacteria* 5 % versus 1.1%. Interestingly, the *Proteobacteria* abundance in *H. pylori* negative stomachs is similar with *Proteobacteria* of *non H. pylori* of the gastric samples, 10.8 ± 3.2 versus 10%.

1.5.2. Bacteria in polluted environments:

Studies in several polluted aquafier habitats revealed different bacterial taxonomic groups depending on the different types of pollutants and environments. For example, members of the genus *Burkholderia* are known to degrade petroleum compounds in the rhizosphere **(115, 116, 117)**. Along with other organisms, such as protozoa, yeasts, unicellular algae and molds, bacteria play an important role in degrading oil polluted marine environments **(118)**. Members of the *Actinobacteria* are known to be responsible for diesel and fuel oil degradation in soil **(119, 120, 121, 122, 123)**. Polycyclic aromatic hydrocarbons (PAHs) are one of the major components in crude oil and bacteria that degrade PAH from soil include members of the *Mycobacterium* genus belonging to the *Actinobacteria*, the genus *Sphingomonas* (α- *Proteobacteria*), the genus *Pseudomonas* (γ- *Proteobacteria*) and the genus *Burkholderia* (β- *Proteobacteria*) **(124)**. Results from the Archipelago Sea (Finland) showed that α- *Proteobacteria* abundance decreased from approximately 54% to 30% due to diesel pollution and yet increased in γ- *Proteobacteria* abundance from 2% to 10%. Diesel fuel polluted sites diesel revealed that members of the *Comamonadaceae* family of the *Burkholderiales* order (β- *Proteobacteria*) accounted for 52% of 16S rDNA clones **(115)**. The primary bacterial genera reported to degrade oil in marine environments includes members of the *Alcanivorax*, *Oleiphilus*, *Oleispira* and *Thalassolituus* **(125)**, *Acinetobacter* and *Pseudomonas* genera **(126)**. All these groups belong to the γ- *Proteobacteria* **(115, 125, 126)**. Other studies showed that γ- *Proteobacteria* members are capable of oil degradation in marine environments **(125, 126, 127, 128, 129, 134)**. *Cyanobacteria* were found to be indirectly involved in oil degradation **(130, 131, 132, 133)** by providing extra O_2 **(133)**. In contrast, other studies of marine ecosystems showed that the dominance of β-*Proteobacteria* and *Actinobacteria* are possibly responsible for diesel degradation **(120, 121, 122, 123, 124).**

A study in urban stormwater sediments in Chassieu, an urban area NE of Lyon, France showed that *Cyanobacteria*, *Bacteroidete*, β-*Proteobacteria*, γ-*Proteobacteria*, α-

Proteobacteria, with proportions from 10%-35% in the community. Moreover, the study revealed that *Nitrospira*, Δ-*Proteobacteria*, *Gemmatimonadetes* are less abundant (<1%) in urban stormwater areas. This urban sediment here was said to be rich in organic compounds, with petroleum by-products which include steranes and terpanes, unresolved complex mixture (UCM) and PAHs **(135).**

Other analyses in streams polluted with high concentrations of uranium, inorganic mercury [Hg(II)], and methylmercury (MeHg), using GS 454 FLX pyrosequencing, revealed the dominance of members of the *Proteobacteria* (ranging from 22.9% to 58.5% per sample), *Cyanobacteria* (0.2% to 32.0%), *Acidobacteria* (1.6% to 30.6%) and *Verrucomicrobia* (3.4% to 31.0%) phyla. Furthermore, some sulfate-reducing bacteria, belonging to the *Proteobacteria* and *Verrucomicrobia* phyla, appeared to be positively associated with Hg and MeHg **(136).**

Analyses of heavy metals in polluted areas in the Xiangjiang river (China) showed that such pollution affected the diversity of the sediment microbial community **(137)**. The dominant phyla in this study were members of the *Proteobacteria,* including α-*Proteobacteria*, β- *Proteobacteria,* and *Firmicutes*. Members of the α-*Proteobacteria* were significantly increased with increasing heavy metal concentrations with proportions ranging from 3.6%-11.5% in less polluted sites to 37.-46.5% in high polluted sites. *Chloroflexi* were found to correlate with high concentrations of Hg in the proportion from 7.6-11.0%, while ε- and α-*Proteobacteria* were positively correlated with Zn and Cd, though Δ-*Proteobacteria* and *Actinobacteria* are negatively correlated with these metals. *Chloroflexi* is positively correlated, while *Firmicutes*, β- and γ-*Proteobacteria* are negatively correlated, with Hg. The study pointed out that ε- and Δ- *Proteobacteria* could be the potential indicators for Zn and Cd contamination of river sediments, and *Chloroflexi* and γ- *Proteobacteria* as indicators for Hg contamination due to their sensitivity to these metals **(137).**

The bacterial composition in a wastewater treatment plant (WWTP) effluent of highly urbanized areas in Chicago was surveyed using tag pyrosequencing of bacterial 16S rRNA genes. The effluence of WWTP had a positive effect on the abundance of *Nitrospirae* and *Sphingobacteriales* and negative effect on the abundance of Δ-*Proteobacteria*, *Desulfococcus*, *Dechloromonas*, and *Chloroflexi* **(138).**

1.6. Project of studying the pollution of the SaiGon-DongNai river system:

1.6.1. The SaiGon-DongNai (SG-DN) river system:

The SaiGon river, which is 225 km in length, originates near Phum Daung (now called Phum Chong Daung) in southeastern Cambodia **(4)**. The river has 250-350m width and 10-20m depth. The SaiGon river is affected by a semi-diurnal tidal flow regime. The DongNai river originates from the Di Linh highland in Vietnam. The total length of DongNai river is 628 km. The two rivers, originating from different areas, join together at Ho Chi Minh City (hereafter HCMC) and finally flow to the East Sea of Vietnam **(5)**.

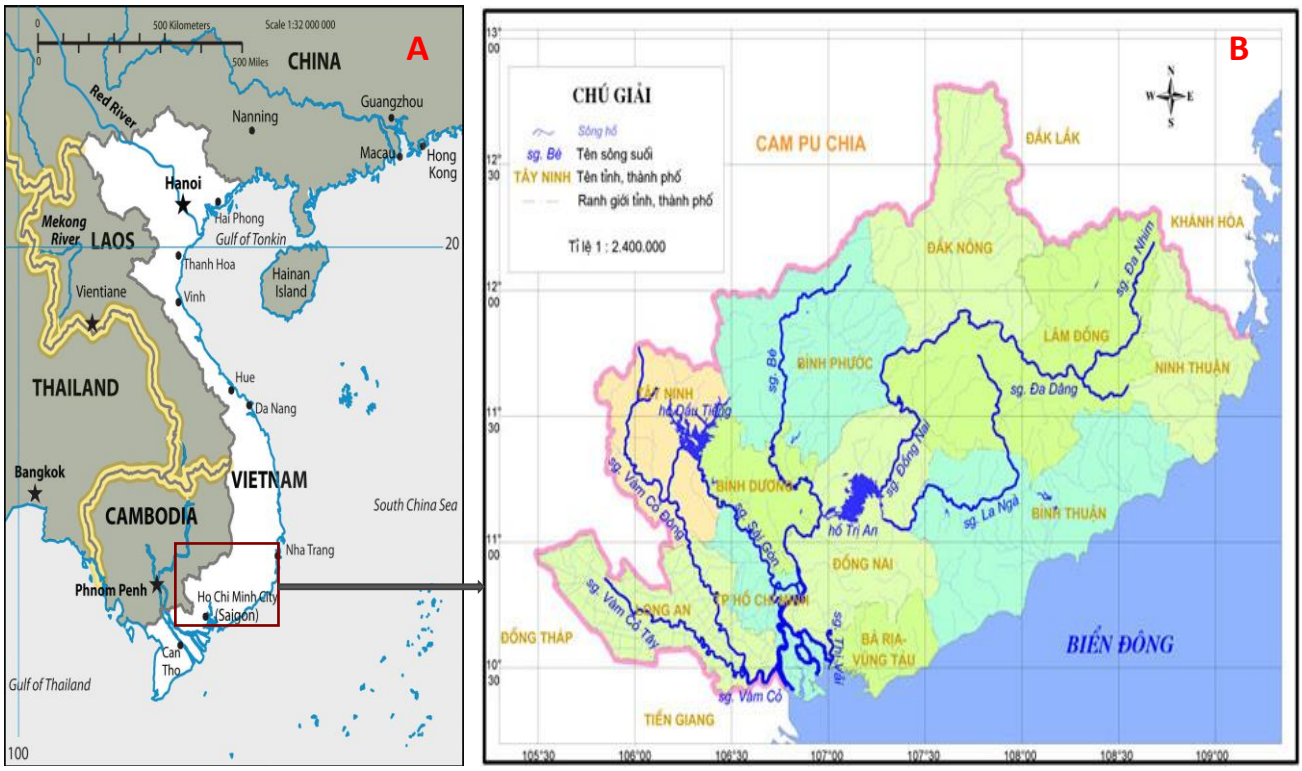


Figure 1.10. The SaiGon – DongNai (SG-DN) river system.

A) Position the SG-DN river system in Vietnam **(6, 7)**.

B) The SG-DN river system running through HCMC and 10 provinces in Vietnam **(8)**.

The SaiGon and the DongNai rivers run through HCMC and 10 provinces, including Dak Nong, Lam Dong, Ninh Thuan, Binh Phuoc, Binh Thuan, Tay Ninh, Binh Duong, Dong Nai, Long An, and Ba Ria-Vung Tau, which play a vital role as the main water resource for local resident activities (**Fig.1.10**). For this reason, the two rivers are called the SaiGon-DongNai river system. The basin of the SaiGon river is about 2,700 km^2 and that of the DongNai river is about 38,610 km^2 **(5)**.

1.6.2. The role of the SG-DN river system in HCMC and 10 provinces:

The SG-DN river system is affected by the population and different activities of the areas. Therefore, information about the population and activities of the areas surrounding is necessary for studying this river system.

1.6.2.1. Population in HCMC and its density:

The population of Vietnam is reported to be 87,840,000 in 2011, representing about 1.28% of the total world population and ranking 14th by country **(9, 10)**. The surface of Vietnam is about 310,070 km^2, ranking 65th of 195 countries in the world. The population densities of Vietnam in 2010 were 268 people per km^2 (people/km^2) and 279 people/km^2 in 2014. HCMC is located in the south of Vietnam and is the largest city in Vietnam with 0.6% coverage of the total area of the country **(5)**.

HCMC is located 1,730 km to the south of Hanoi, the capital of Vietnam, and also at the crossroads of international maritime routes **(5)**. It is a transport hub of the southern region with the largest port system and airport in Vietnam.

The population of HCMC is 7,521,138, ranking in 1st position of the total national population, with the population density of 3,590 /km^2, accounting for about 8.6 % of the total country population (data from 2011) **(11)**. However, the population density of HCMC is different if counting the whole city or just its urban districts. The density of the urban districts is about 12,449 people/km^2 **(11)**. By comparison, the densest urban area in the world is Dhaka (Bangladesh) with 35,000 people/ km^2 in the year 2010 **(12)**.

1.6.2.2. Other 10 provinces population and their densities:

The SG-DN system runs through 10 provinces and HCMC (**Fig. 1.10 B**), which have a combined population of 19,824,700 (**Table 1.1**). Among these, Dong Nai and Binh Duong provinces had the highest population and density with 2,665,100 and 1,691,400, 451 people/km^2 and 628 people/km^2, respectively.

1.6.2.3. The role of the SG-DN river system in HCMC and 10 provinces:

The SG-DN river system is the main water resource for about 19 million residents of HCMC and 10 provinces, though there are other water resources, such as groundwater and rain water, with groundwater accounting for 30-40% of water demand in HCMC **(5).**

According to the Saigon Water Supply Company (SAWACO), the SG-DN river system provides 1,150,000 m^3 of water per day (m^3/day) out of a total of 1,236,000 m^3/day for HCMC in 2006. Besides HCMC, other provinces such as Dong Nai, Binh Duong, Ba Ria-Vung Tau, Tay Ninh and Long An also use the water from the SG-DN river system as

the main water source. The intake rate of these provinces is from 3,700 m^3/day to 100,000 m^3/day according to data from 2005 **(5)**. Moreover, there are 157 cooperatives, 12 craft villages, 43,000 enterprises and around 60 industrial zones located around the river **(14)**. In HCMC, there are about 30,000 factories, including many large and small enterprises, high technology, electronic, processing, construction, building materials and agro-products, plus 15 industrial parks (IP) and export-processing zones (EPZ). There are 171 medium and large-scale markets, tens of supermarket chains, dozens of luxury shopping malls and many modern fashion and beauty centers **(5)**. The river also provides fresh water for 1.8 million hectares of cultivated land around HCMC and 11 surrounding provinces **(15)**.

Table 1.1: Population, density and area of the 10 provinces and HCMC in 2011 **(13)**.

Provinces & HCMC	Population	Population density (people/km^2)	Area (km^2)
Dak Nong	516,300	79	6,515.6
Lam Dong	1,218,700	125	9,773.50
Ninh Thuan	569,000	169	3,358.30
Binh Thuan	1,180,300	151	7,812.90
Binh Phuoc	905,300	132	6,871.50
Dong Nai	2,665,100	451	5,907.20
Tay Ninh	1,080,700	268	4,039.70
Binh Duong	1,691,400	628	2,694.40
Long An	1,449,600	323	4,492.40
Ba Ria – Vung Tau	1,027,200	516	1,989.50
HCMC	7,521,100	3589	2,095.60
Total	19,824,700	356.9	55,550.60

1.6.3. Introduction of pollution in the SG – DN river system:

1.6.3.1. Continuously national reports of the pollution in the SG – DN river system:

The SG-DN river system provides fresh water for daily activities of local resident such as drinking, cooking, bathing, for agriculture and also for transportation purposes.

The SG-DN river system not only supplies the water for surrounding citizens but also carries away much daily waste **(68)**. Furthermore, the river also conveys the waste of

surrounding industrial park, hospitals, as well as agricultural run-off **(68)**. It is reported that the daily volumes of domestic and industrial wastewater discharged to the canals in HCMC in 2000 were 710,000 m^3 and 35,000 m^3, respectively **(5, 68)**. However, currently, only 4% (about 30,000 m^3 /day) of the municipal wastewater was conventionally treated at the Binh Hung Hoa central wastewater treatment plant **(5, 68)**. For industrial waste water, about 40% (approximately 15,000 m^3/day) of wastewater was treated by the centralized wastewater treatment plants located inside the five industrial parks (the Tan Thuan, Linh Trung 1, Linh Trung 2, Tan Binh, Le Minh Xuan, and Tan Tao), leaving 60% (about 20,000 m^3) untreated **(5)**. This was confirmed by the Ministry of Science and Technology on Thanhnien News on October 13, 2013, which stated that 66 % of 179 operating industrial zones were using or building wastewater treatment plants, and only around 58 % of the daily discharge of 622,773 m^3 was treated before reaching the waterways **(14)**. On December 13, 2009, the Ministry of Natural Resources and Environment reported that the Dong Nai river received about 1.54 billion liters of wastewater from 70 industrial parks per day, together with 1.73 billion liters of wastewater from residential areas **(69)**. Moreover, in 2008 Dr. Vo Le Phu reported that about 17,000 m^3 of hospital effluents were discharged into the SG - DN river daily. In the same report, the author also said that, according to the Environmental Management Division of Department of Natural Resources & Environment (DONRE), only 40% of this wastewater is treated **(68)**. In addition, the SG-DN catchment and watershed also receives agricultural run-off from high usage of chemical, organic fertilizers and pesticides **(70)**.

The discharge of wastewater from domestic, industrial and agricultural activities into the SG-DN river system has been reported to be responsible for pollution and affecting aquafier life **(5)**. The online News Vietnam Plus, in April 2010, published that ‘‘The water in some downstream sections of the DongNai river system had become badly polluted, exceeding danger levels, according to the Pollution Control Department (PCD) under the Ministry of Natural Resources and Environment ’’ **(71)**. In early April 2000, more than 50 tons of fish died in an upstream feeder of the Tri An Reservoir in the upper reaches of the DongNai river **(70)**.

In addition, biological pollution related to the SG-DN river has also been reported. Dozens of water samples failed safety standard tests in HCMC in March 2009. Mr. Le Truong Giang, deputy director of the city's health department said “We detected bacteria in our samples, mainly *Coliform* and *Pseudomonas aeruginosa*," leading to more than 38 water bottling firms being ordered to close **(72)**. A test conducted in the fourth

quarter of 2009 (from September to December) by the Dong Nai Department of Natural Resources and Environment found 8 potentially harmful elements exceeding safe levels in the DongNai river, including *Coliform* bacteria and total suspended solids (TSS) **(69)**. The tested samples were collected in a river section in the town of Bien Hoa (locating along the DongNai river), which is also the source of more than one billion liters of tap water pumped to HCMC every day **(69)**. In the same river section, the HCMC Preventive Health Center also found high concentrations of organic substances and iron. In October 2009, the agency also took samples at the source for the Binh An Water Supply Company **(73)** and found concentrations of 1.38 mg of iron and 0.8 mg of ammonium ions per liter of water. The allowed levels per liter are 1 mg of iron and 0.2 mg of ammonium ions **(69)**.

Mr. Nguyen Hoang Hung, director of the Dong Nai Department of Natural Resources and Environment, Environmental Monitoring Center, said that the DongNai river section in Bien Hoa was suffering from uncontrolled wastewater discharges from industrial parks. The center also reported that four industrial parks along the DongNai river were discharging untreated wastewater because they did not have any wastewater treatment systems, these include 100 firms at the Bien Hoa 1 Industrial Park (IP) and Mr. Hoang Van Thong, head of the Dong Nai Environment Protection Agency, said most of these 100 facilities were built in the 1970s without any wastewater treatment plants. The Bien Hoa 1 IP asked the Bien Hoa 2 IP to help treat 600,000 liters of the daily 15 million liter-wastewater effluent, leaving the rest go directly into the DongNai river. Professor Lam Minh Triet, an expert of the DongNai river, at the Vietnam National University (in HCMC), said that many provinces, like Binh Duong and Dong Nai, had focused solely on setting up IPs rather than on waste control or environmental protection. Scientists have been warning of pollution issues in the SG-DN river system for ten years, but every warning has been ignored **(69)**.

1.6.3.2. Pollution of the urban canal systems:

More than half of the canal systems in HCMC represent high-density population and industrially polluted areas. Unfortunately, domestic and industrial wastewater is directly released into the water and canal systems without, or with inadequate, treatment.

The Tan Ho-Lo Gom, Nhieu Loc-Thi Nghe, Tau Hu-Ben Nghe and Kenh Doi-Kenh Te canals receive about 700,000 m^3 of municipal and industrially effluent with high levels of BOD, COD and heavy metals, well above Vietnamese standard levels **(68)**. DONRE noted that all HCMC's rivers and canals were heavily polluted by organic wastes and *Coliform* bacteria, particularly in the dry season **(68)**. Furthermore, the canal systems

situation is increasingly aggravated during wet weather, as the canals receive additional contaminated flows from urban and agricultural runoff. Not surprisingly, high concentrations of PCBs, DDT and heavy metals were found in canal sediments **(68)**.

1.6.3.3. Evidence of untreated wastewater from industrial factories polluted the river:

1.6.3.3.1. Scandal of Thi Vai river pollution:

The rising level of pollution in Vietnam's waterways has been common knowledge for years. But, in October 2008, it became a public scandal after Vedan, a Taiwanese maker of monosodium glutamate (MSG), confessed to discharging toxic waste through hidden pipes into the river for years **(74)**. Vedan Vietnam was caught in 2008 discharging untreated effluents directly into the Thi Vai river in Dong Nai province through secret pipes (**Fig. 1.11**, **1.12**) **(75, 76, 77, 78)**. It had been doing this for 14 years, from 1994 to 2008, inflicting serious damage on the river system, fish farms as well as rice fields located on the banks of the river **(74, 79)**. The factory was found to have discharged between 35,000 and 45,000 m^3 of untreated wastewater directly into the Thi Vai river every day for a decade and a half, which led to the damage of nearly 2,000 hectares of fish and shrimp ponds in Dong Nai, Ba Ria – Vung Tau provinces **(69)**. However, Vedan has refused to take full responsibility. The company said that the study was made in the dry season when high salinity levels could affect the results **(69)**. In addition, there were many complaints from the residents living around the river with residents along the Thi Vai complaining about the critical situation of the waterway in the river for over a decade''**(80)**.

Figure 1.11. Location of Vedan company on the Thi Vai river **(76)**. Note: red mark is location of Vedan company.

Figure 1.12. Evidence that the Vedan factory dumped waste into the Thi Vai River of Dong Nai Province (photo courtesy of Tuoi Tre) **(77, 78)**.

1.6.3.3.2. Vedan admits to polluting parts of Thi Vai River:

Finally, on Vietnam News, in December 2009, the Vedan company admitted to discharging untreated wastewater into the Thi Vai River, which polluted an 11-kilometre stretch of the waterway in southern Dong Nai province. Vedan general director Yang Kun Hsiang was quoted by Nguoi Lao Dong (The Labourer) newspaper saying that the company took responsibility for only 60 to 70 % of the river's pollution. However, according to the Institute of Environment and Natural Resources, which belongs to the

National University in HCMC, Vedan's waste proportion was 89-98% based on analysis of the pollution source. This was confirmed by professor Le Quoc Hung, from the Viet Nam Institute of Sciences and Technology, showed that water quality examined in November in the river had improved after Vedan stopped its release early in 2009 **(74, 82)**.

Additional evidence of untreated industrial waste water released into the environment were the reports in 2001 of companies such as the Phuoc Long Textile company, Cofidec, a seafood processing company, and the Mai Tan Paper Company discharging about 1500 m^3, 90 m^3 and 300 m^3 of untreated wastewater, respectively, into the water on a basic daily **(80)**.

1.6.3.4. The impact of pollution on the living condition of surrounding habitats:

The growing population of HCMC and provinces such as Dong Nai and Binh Duong leads to increased pollution of the SG-DN river system and pressure on the fresh water supply for surrounding residents **(5, 68)**. In 2009, the vice director of the Thu Duc Water Supply Company in HCMC, Mr. Truong Khac Hoanh, warned that with such an increase in pollution, this major water supply would soon be not usable and aquatic life would not be able to survive due to high levels of pollution **(69)**. The southeast river cluster of the DongNai river is expected to be at risk of exceeding project water needs in 2020 **(82)**. Therefore, there have been numbers of national and international scientific studies published to show the extent pollution of the SG-DN river system.

1.6.4. The goal of my project:

The goal of my project includes two main parts:

First, the chemicals of the SG-DN river sediment were analyzed at two different time points, February 2012 and August 2012, in order to evaluate the pollution of the SG-DN river based on its highly industrial and urban activities.

Second, to attempt to correlate the impact of industrial and urban activities on the bacterial community of the river sediments based on the identification of V3-V1 sequences of the 16S rDNA gene at the phyla, genus and OTUs level.

CHAPTER 2: MATERIALS & METHODS

2.1. Studying sites & Sampling method:

2.1.1. Map of studying sites:

The SaiGon-DongNai river system contains 2 rivers: the SaiGon and DongNai rivers. They join at HCMC before flowing to the sea. The part of the river system which goes through Binh Duong, Dong Nai provinces and HCMC were of the most concern here, due to their high population and dense industrial zones (**Fig.2.1, 2.2**) There have been continuous reports by local and national media about the severe pollution levels in this part of the river system (section 1.6.3.1 of Introduction chapter).

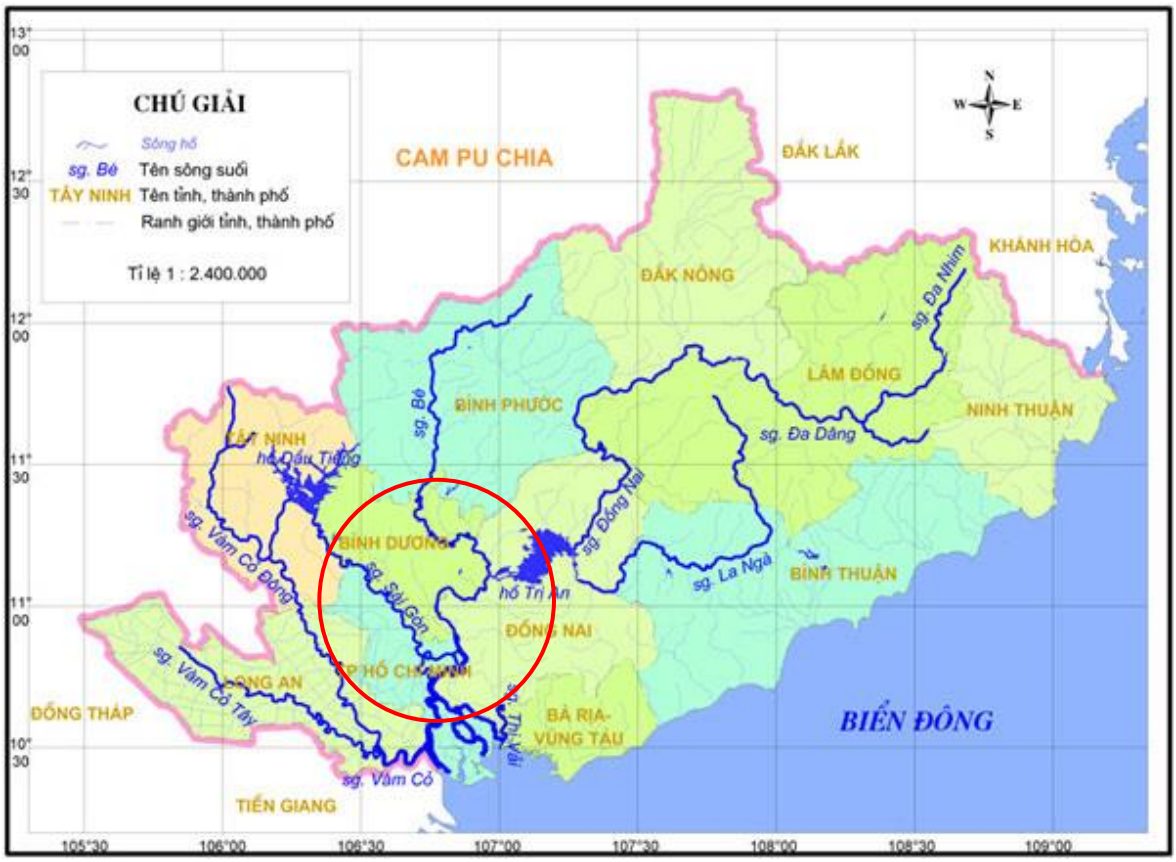


Figure 2.1. The studied area in the SaiGon-DongNai river system. Note: the studied locations in this project are marked in red circle **(8)**.

Thirteen locations of the SaiGon – DongNai river system were chosen according to the Institute of Environment and Resources's "pollution potential" documents and previous studies (**Fig.2.1**) **(235, 236)** with (i) 5 locations belonging to the SaiGon branch from upstream to downstream which are called location Reference 1, SaiGon 1, SaiGon 2, SaiGon 3 and SaiGon 6 (abbreviation are RF1, SG1, SG2, SG3 and SG6 subsequently); (ii) 5 locations belonging to the DongNai river from upstream to downstream which are called location Reference 2, DongNai 1, DongNai 2, DongNai 3 and SaiGon 4 (RF2, DN1, DN2, DN3 and SG4); (iii) 1 location belonging to the intersection between the SaiGon and the

DongNai rivers called SaiGon 5 (SG5) and (iv) 2 locations belonging to the canals of the SaiGon – DongNai river basin are SaiGon 8 and SaiGon 9 (SG8 and SG9) (**Fig.2.2**). The latitude and longitude of the 13 locations is shown in **Table 2.1**.

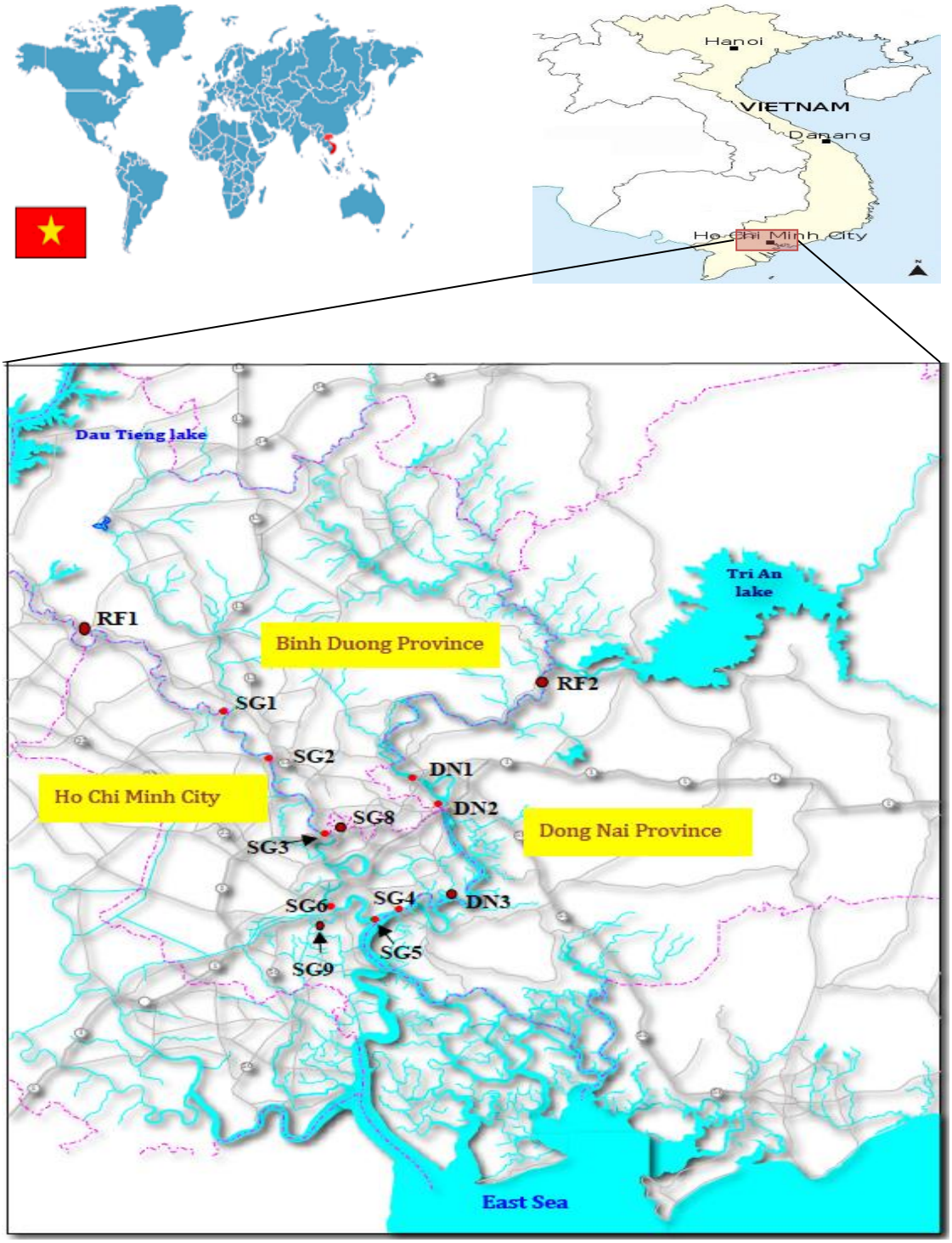


Figure 2.2. The map of thirteen studying locations.

2.1.2. Sampling method:

Sediments were sampled using an Eckman grab near the bank of the river with water depth ranging from 1 to 3 meters **(235)**. Surface sediments were collected from the top 1-10 cm. For each location, the samples were taken on the left side and the right side of the river, called the A side and the B side, respectively **(236)**. For the biogeological replication, 2-3 sediment samples were collected for each A and B side (**Fig. 2.3**). Due to difficulties of transportation, the team was not able to collect the 6 biological replicates for each location. Total 42 sediment samples according to 13 locations were collected (**Fig. 2.4**). Each sediment sample was placed into a sterilized plastic box and stored at 4^0C during transport **(235, 236)**. Sediment samples were obtained in August 2012 with PhD. Nguyen Ngoc Vinh (Institute for Environment and Resources, Vietnam National University of Ho Chi Minh City (VNU-HCM)) **(235, 236)**.

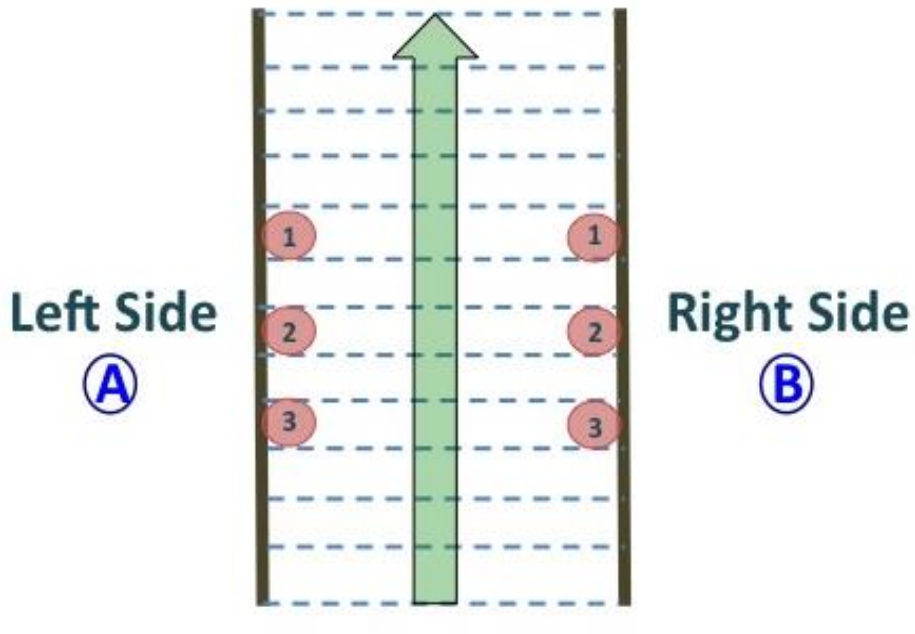


Figure 2.3. Sampling method. For each location, the samples were taken at the left side and the right side of the river called the A side and the B side subsequently. For the biogeological replication, 2-3 sediment samples were collected for each A and B side. Note: the green arrow represents the direction of river flow.

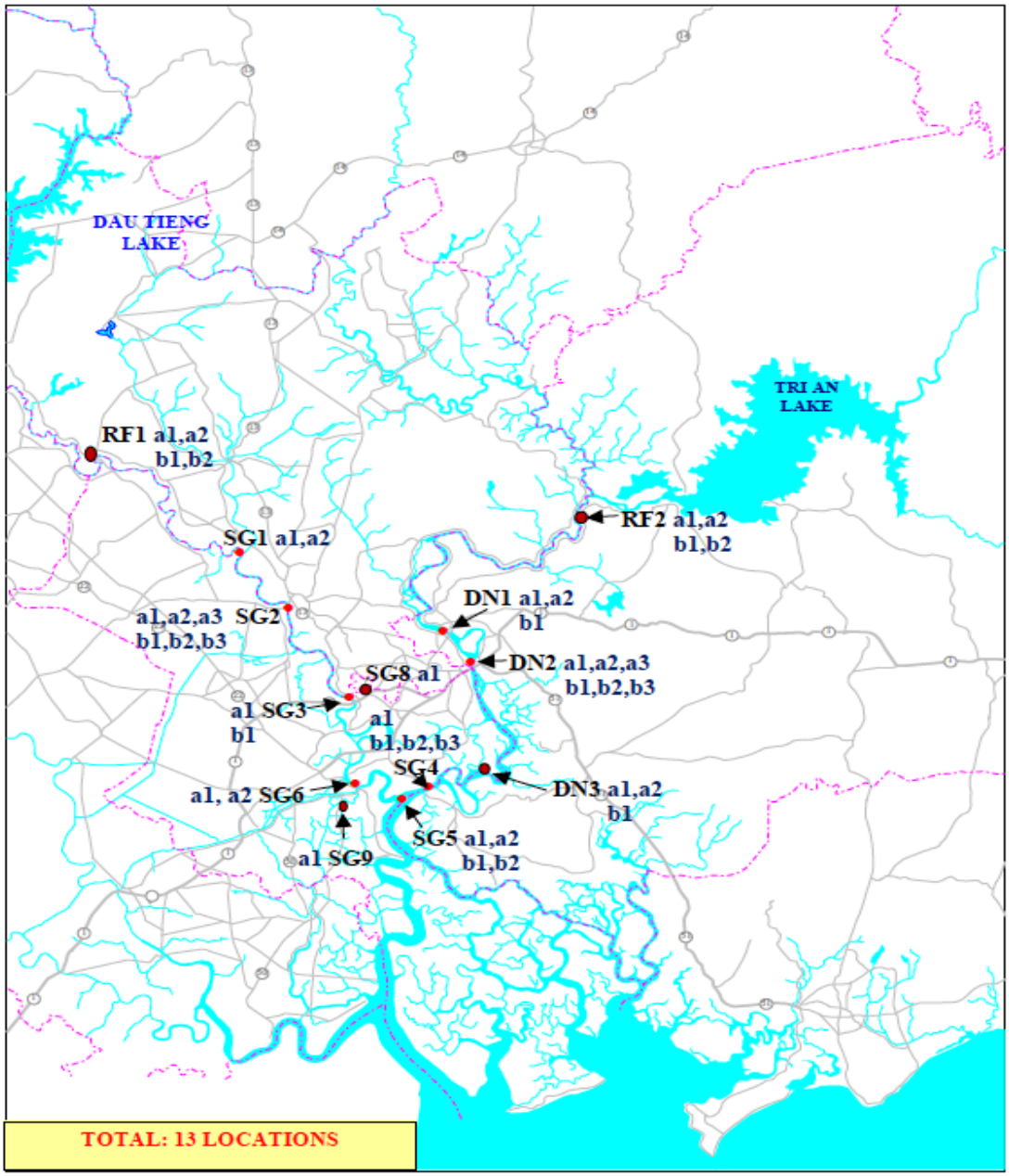


Figure 2.4. Map of 42 collected sediment samples from 13 locations. Note: For the biogeological replication, 2-3 sediment samples were collected for each A and B side. Due to difficulties of transportation, the team could not collect 6 biological replicates for each location.

Table 2.1: Latitude and longitude of sediment samples from the thirteen locations.

Location		Latitude	Longitude
SaiGon river	RF1a	N 11^0 9’ 19,44”	E 106^0 27’ 4,48”
	RF1b	N 11^0 9’ 18,99”	E 106^0 27’ 6,92”
	SG1a	N 11^0 02’ 28,46”	E 106^0 36’ 18,09”
	SG2a	N 10^0 59’ 8,16”	E 106^0 37’ 20,47”
	SG2b	N 10^0 59’ 5,36”	E 106^0 37’14,02”
	SG3a	N 10^0 51’ 42,63”	E 106^0 43’ 4,74”
	SG3b	N 10^0 51’ 39,37”	E 106^0 42’ 58,71”
	SG6a	N 10^0 45’ 38,26”	E 106^0 43’ 21,89”
DongNai river	RF2a	N 11^0 4’ 7,04”	E 106^0 57’ 2,93”
	RF2b	N 11^0 4’ 10,50”	E 106^0 56’ 59,90”
	DN1a	N 10^0 57’ 0,67”	E 106^0 48’ 21,95”
	DN1b	N 10^0 56’ 43,09”	E 106^0 48’ 16,14”
	DN2a	N 10^0 54’ 28,39”	E 106^0 50’ 29,20”
	DN2b	N 10^0 54’ 15,11”	E 106^0 50’ 12,21”
	DN3a	N 10^0 46’ 38,73”	E 106^0 51’ 17,55”
	DN3b	N 10^0 46’ 38,08”	E 106^0 51’ 16,09”
	SG4a	N 10^0 45’ 6,59”	E 106^0 47’ 24,46”
	SG4b	N 10^0 45’ 25,49”	E 106^0 47’ 15,44”
Intersection	SG5a	N 10^0 44’ 15,38”	E 106^0 46’ 18,93”
	SG5b	N 10^0 44’ 52,18”	E 106^0 45’ 47,57”
Canals	SG8a	N 10^0 53’ 15,12”	E 106^0 43’ 44,03”
	SG9a	N 10^0 45’ 9,76”	E 106^0 42’ 0,45”

2.2. DNA extraction and PCR for pyrosequencing:

Total DNA from sediment samples were extracted using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc, CA, USA). Then, PCR reactions were performed using the universal 16S rDNA bacterial primers 27F(*A*AGAGTTTGATCMTGGCTCAG) and 518R(*B*xxxxxxxxxxWTTACCG-CGGCTGCTGG) where *A* and *B* represent the adaptors A and B for pyrosequencing using the GS Junior Titanium emPCR Kit (Lib-A) reaction (GS Junior, Roche/454 Life Sciences, Branford, CT, USA). The xxxxxxxxx represents ten nucleotide sequence tags designed for sample identification barcoding **(237, 238).**

Total DNA for each sediment sample was subjected to three PCR reactions per thermostable DNA polymerase, and two different thermostable DNA polymerases were used for each sample to reduce potential PCR bias.

PCR amplification conditions were modified for the use of two different thermostable DNA polymerases per sample: (I) Phusion High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland)— 98 °C for 2 min followed by 30 cycles of 98 °C for 30 s, 56 °C for 20 s and 72 °C for 20 s and a final elongation step at 72 °C for 10 min. (II) High Fidelity PCR Enzyme Mix (Fermentas)—94 °C for 3 min followed by 30 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 40 s, and a final elongation step at 72 °C for 10 min.

Each 25 μl volume PCR reaction was performed with 1–10 ng DNA template, 0.5 μM of each primer (Sigma-Aldrich, MO, USA), 0.2 mM dNTP mix (Fermentas), 0,025 units of Phusion-High Fidelity (I) or 0,625 units of High Fidelity PCR enzymes (II). A total of six PCR reactions were obtained for each DNA sediment sample.

The 5μL of six PCR products for each DNA sediment sample were pooled together and loaded on a 0.8 % agarose gel in 1X TAE buffer. After electrophoresis and DNA visualization by ethidium bromide staining and long wave UV light illumination (365 nm), the 500-600 bp fragments of amplified 16S rDNA were cut from the gel and purified using the NucleoSpin Extract II kit (Macherey-Nagel, North Rhine-Westphalia, Germany) according to the manufacturers' instructions. Pooled PCR products of each DNA sediment sample were adjusted to 56 nanograms and mixed together for pyrosequencing **(239)**.

2.3. Pyrosequencing:

PCR products were sent to the Department of Biology (University of Oulu, Finland) for pyrosequencing using the GS Junior system.

2.4. Sequencing data processing pipeline:

After sequencing, the sequences were first selected by their length (200-600 bases) and containing homopolymers ≤ 8nt in length. Then, sequences were retained with zero error in the barcoding tags and ≤ 1 error in the 5' primer, performed with the Mothur pipeline **(240, 241)**. The 3'adaptors and 3'primer sequences remaining from the sequencing process were removed using the Cutadapt tool implemented on the Galaxy server of the Institut de Génétique et Microbiologie (IGM) of the Université Paris-Sud (http://galaxy.igmors.u-psud.fr/) **(242)**. Sequences were then quality trimmed using Condetri V_2.2 (243) and the adjusted parameters were: consecutive high quality bases (n_H=10), consecutive low quality bases (n_L=1), threshold high quality score (Q_H=25), threshold low quality score (Q_L =10), the fraction f of bases with a quality score higher than Q_H is 80%, and the fraction f of bases with a quality score less than a lower bound threshold Q_L=10 is 0%. Sequences containing more than one ambiguous base (N) were removed using Mothur. Chimeric sequences were removed using the Decipher (244) and Uchime **(245)** programs together. Uchime program were adjusted by (i) replacing its database by modified Greengenes database **(246)** and (ii) adjusted parameter mindiv=1.5. The modified Greengenes database is the 97_OTUs fasta file (from May 2013) downloading from Greengenes website (http://greengenes.secondgenome.com/downloads) and then run through Uchime program (using parameter mindiv=1.5) with itself as the reference database. The 97_OTUs fasta file is the whole Greengene database that group into OTUs with 97% similarity. Sequences of each sample were normalized to 2983 reads by random selection using the function of rarefication_even_deepness without replacement of the phyloseq library from R **(247)**.

The sequences were assigned into Operational Taxonomic Units (OTUs) using the CD-HIT-OTU method (**248**) with a 97% threshold; and individually classified using the Silva NGS website with Silva database release 123 **(249, 250)**, with 100% sequence similarity for clustering and 90% as a classification similarity threshold.

`

2.5. Statistical analyses:

All statistical analyses were conducted using R version 3.1.2 (R core Team, 2014) with 0.05 as the significance threshold. Principal component analyses (PCA) were performed first (i) on the relative proportion of phyla or genus with each and all PAH compounds, second (ii) among the samples using the ade-4 package adapted in R **(251)**. The diversity and richness indices (Chao1 and Shannon) of the samples were estimated and the distances among the samples were calculated based on Bray-Curtis metrics with similarity index (97% cutoff), using the package Phyloseq **(247)**. Then, the unweighted pair group method with arithmetic mean (UPGMA) clustering tree was built based on Bray-Curtis metrics using the ade-4 package **(251)**. All sequences have been deposited in the GenBank Sequence Read Archive with accesion numbers SRP090995. The pipeline is summarized in **Fig. 2.5**.

2.6. 16S copy numbers normalization:

High-quality sequence after going through the cleaning process from step 1 to step 7 (**Fig.2.5)** were normalized for the 16S copy numbers by the Tax4Fun **(346)**.

For the normalization of 16S rDNA copy number by Tax4Fun program. The SILVA-based 16S rRNA profile is transformed to a taxonomic profile of the prokaryotic KEGG organisms. Then, the estimated abundances of KEGG organisms are normalized by the 16S rRNA copy number obtained from the NCBI genome annotation.

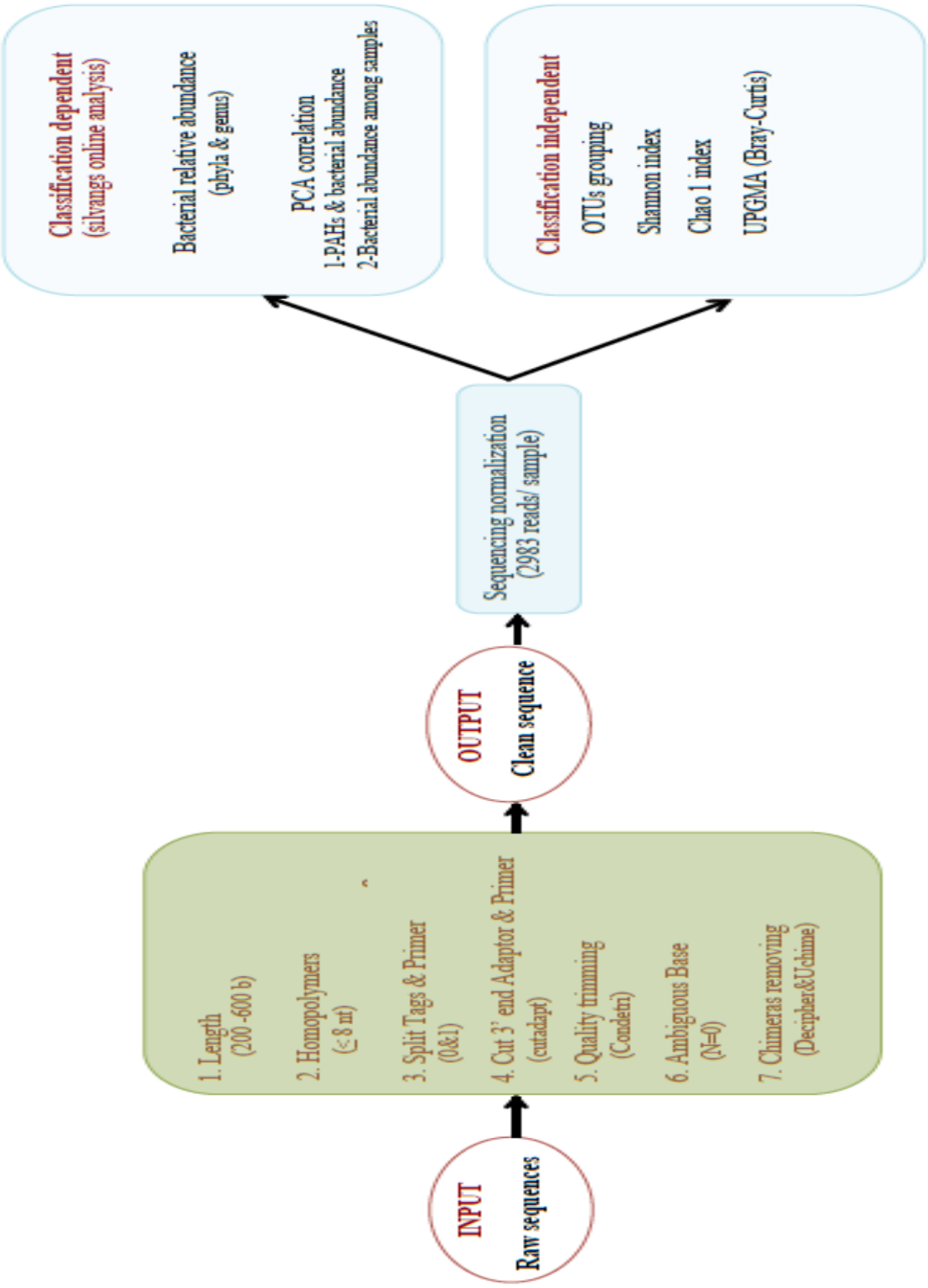


Figure 2.5. Sequence processing pipeline and data analyses.

2.7. Searching for chemical and biological pollutants:

2.7.1. Data obtained in February 2012:

The SG-DN sediment samples were collected on February 2012 for chemical analyses in order to evaluate the pollution levels of the river. Due to financial limitations, only sediment samples from 8 out of 13 locations were collected, including 4 locations of the SaiGon river : SG1, SG2, SG3, SG6 and 3 locations from the DongNai river : DN1, DN2, SG4 and the intersection location of the two rivers SG5 (**Table 2.2**).

For chemical analyses, the sediment samples were taken at two sides of the river and then mixed together for each location. Samples were stored in sterile plastic tubes and covered by aluminum foil, at 4^oC. The date of sampling was 29^{th} February 2012 by Dr. Nguyen Ngoc Vinh in Ho Chi Minh Science and Technology Institute (HCMC, Vietnam) (**235**).

Table 2.2: Latitude and longitude of sediment samples from 8 locations taken in February 2012.

Location		Location name	Latitude	Longitude
SaiGon river	SG1	Thi Tinh River	N 11^0 02’ 24,75”	E 106^0 36’ 10,05”
	SG2	Hoa Phu Pumping Water Station	N 10^0 59’ 6,74”	E 106^0 43’4,5”
	SG3	Binh Phuoc Bridge	N 10^0 51’ 42,84”	E 106^0 43’4,5”
	SG6	Tan Thuan Bridge	N 10^0 44’ 11,38”	E 106^0 46’ 15,04”
DongNai river	DN1	Hoa An Bridge	N 10^0 57’ 16”	E 106^0 48’ 20,86”
	DN2	Dong Nai Bridge	N 10^0 54’ 25,06”	E 106^0 50’ 26,83”
	SG4	Cat Lat Port	N 10^0 45’ 05,82”	E 106^0 47’ 23,27”
Intersection	SG5	Mui Den Do	N 10^0 44’ 11,38”	E 106^0 46’ 15,04”

2.7.1.1. Total organic carbon (TOC):

Total organic carbon was measured in the laboratory of Institute for Environment and Resources, Vietnam National University of Ho Chi Minh City (VNU-HCM)] (**235**, **236**).

2.7.1.2. Heavy metals:

Seven heavy metals (Hg, Pb, Cd, Cu, Ni, Cr and Zn) of sediment samples were analyzed in Hoan Vu Hoan Vu Scientist Technologies Company Limited (Ho Chi Minh City, Vietnam) (Annex 5) with the methods described in **Table 2.3**.

Table 2.3: Method of each chemical analyses.

Chemical	Method
Hg	US EPA SW 846 Method 3050 B &SMEWW 3120 B – IC (256-7) [a]
Pb	US EPA SW 846 Method 3050 B
Cd	US EPA SW 846 Method 3050 B &SMEWW 3120 B – IC [b]
Cu	US EPA SW 846 Method 3050 B
Ni	US EPA SW 846 Method 3050 B
Cr	US EPA SW 846 Method 3050 B
Zn	US EPA SW 846 Method 3050 B

Note :

Units of all chemical analyses are $mg.kg^{-1}$

a : detection threshold < 0.001 $mg.kg^{-1}$

b : detection threshold < 0.001 $mg.kg^{-1}$

2.7.1.3. Polycyclic aromatic hydrocarbons (PAHs):

A total of 13 standard PAHs compounds (**Table 3.27**) were detected by the PP AOAC 2007-01 method performing at the Hoan Vu Scientist Technologies Company Limited (Ho Chi Minh City, Vietnam) (**Annex 5**).

2.7.1.4. Polychlorinated biphenyls (PCBs):

PCBs was measured by GC/MS (gas chromatography/mass spectrometry) in the laboratory of the Institute for Environment and Resources, Vietnam National University of Ho Chi Minh City (VNU-HCM) by PhD. Nguyen Ngoc Vinh (**235**, **236**).

2.7.2. Data obtained in August 2012:

Due to the financial limitations, only one chemical pollutant (PAHs) and biological "pollutants" (*Fecal Coliforms* and *Escherichia coli)* were analyzed for the sediment samples collected in August 2012. The analyses were performed for the samples of the left and right sides of each location, called a1 & b1. There are a total of 22 samples (of 42 total sediment samples) that were analyzed for chemical and biological values.

2.7.2.1. PAHs:

A total of 17 standard PAH compounds (compared to 13 compounds analyzed on February 2012) were also detected using the PP AOAC 2007-01 method performing at the Hoan Vu Scientist Technologies Company Limited (HCMC, Vietnam) (Annex 5). Four additional PAH compounds that were analyzed include fluorene, perylene, benzo(j)fluoranthene and benzo(e)pyerene (**Table 3.28**).

2.7.2.2. Searching for Fecal Coliforms and Escherichia coli:

2.7.2.2.1. Searching for Fecal Coliforms:

First, sediment samples were gently mixed with sterilized glass sticks. Then, 1 g of well-mixed sample was homogenized in 100 ml NaCl (0.9%) shaking for 10 mins. Homogenized samples were diluted as 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}. Diluted sample were incubated in a waterbath at 44.5°C for 24 ± hours with Lauryl Trypton Broth media. After incubation, the samples were examined for growth and gas production (gas production with growth is considered a positive fecal coliform reaction). This method was performed according to Method 1681: Fecal Coliforms in Sewage Sludge (Biosolids) by MultipleTube Fermentation using A-1 medium July (2006) (**252**).

2.7.2.2.2. Fluorogenic detection of Escherichia coli:

From each gassing fecal coliform tube, samples were streaked for isolation on a 96-well plate with EC- MUG (EC medium with 4-methylumbelliferyl-β-D-glucuronide) agar and incubated overnight at 37ºC. *E. coli* produces the enzyme glucuronidase that hydrolyzes MUG to yield a fluorogenic product detectable under long-wave (366 nm) UV light (253). After incubation, the 96-well plate was observed under the UV light. Wells that emitted the fluorescent light were considered positive. Most probably number (MPN) of *E .coli* is calculated based on the method described (**254**).

CHAPTER 3: RESULTS

3.1. Test of bioinformatics tools:

3.1.1. 16S rDNA universal primers evaluation:

3.1.1.1. Primers collection:

Criteria for selecting 16s rRNA primers for amplicon sequencing:

i. The amplicon has to be compatible with the sequencing strategies which depend on the read length generated by the sequencing platform and single-end or pair-end sequencing type.
ii. The amplicon is capable of identifying, as much as possible, bacteria which are present in the samples (environmental samples such as soil, sediment, water, biofilm, or clinical samples).
iii. The amplicon is suitable for taxonomic classifier.

For amplicon to be capable of identifying bacterial species or genera in the samples:

- The primer-pair has to be in conserved regions of the 16s rRNA gene.
- The forward and reverse primers match closely to the sequences present in the 16S rDNA gene databases.

In order to define good primers for identifying the community in the environmental samples, 16 primer pairs of 16S rDNA gene from 10 articles published between 2008-2013 were collected and presented in **Table 3.1**. Then, the coverage percentage of these primers in 16S rDNA databases was examined.

Table 3.1: Names and sequences of 16 primer pairs.

Nº	Forward primer	Reverse primer	Region covered	Amplicon length (including primer pairs)	Technique used	References
1	27F	515R	V1-V3	526	454 FLX Titanium	**325**
2	27F	518R	V1-V3	527	454 FLX Titanium	**326**
3	27F	518R	V1-V3	526	454 FLX Titanium	Lab's primer
4	357F	785R	V3-V4	465	454 FLX Titanium	**327**
5	357F-1	909R	V3-V5	586	454 FLX Titanium	**328**
6	355F-2		V3-V5	572		
7	355F-3		V3-V5	571		
8	937F	1492R	V6-V9	587	454 FLX Titanium	**330**
9	357F	785R	V3-V4	449	In-silico/454 FLX Titanium	**331**
10	357F	515R	V3	177	Roche GS FLX	**332**
11	27F	515 R	V1-V3	522	454 FLX Titanium	**333**
12	531F	1100R	V4-V5	584		
13	27F	518R	V1-V3	527	454 FLX Titanium	**334**
14	357F	909R	V3-V5	586		
15	984F	1492R	V6-V9	545		
16	534F	786R	V4	291	Illumia	**329**

Note:

The 1st column is the collected 16S primer pairs in the order from Nº 1 to Nº 16.

The 2nd column is the names of the 16S forward primers, for example, 27F.

The 3rd column is the names of the 16S reverse primers, for example, 515R.

The 4th column is the regions covered by the 16S primer pairs.

The 5th column is the amplicon length that were generated by the primer pairs.

The 6th column is the techniques that used for amplifying the amplicon.

The 7th column is the publications that the 16S primer pairs were collected.

Forward primers from N^0 1 to N^0 16 were organized according to the differences of their sequences, length and positions on the 16S rDNA gene (according to *E. coli* positions) with the regions from V1 to V6 (**Table 3.2**).

Table 3.2: Similarities and differences between 16 forward primers. Among all 16 forward primers collected, primers N^0 1, 11, 13 are identical. Primers N^0 2 & 3 are similar with N^0 1, 11, 13. Primers N^0 4 & 10 are identical. The other primers are different in the ambiguous nucleotides (e.g. M or W letter), their position in the 16S rDNA gene, and their length.

N^0	Name	5'-->3' sequence	Position [a]	Length	Region started [b]
1	27F	AGA GTT TGA TCC TGG CTC AG	8-27	20	V1
11	27F	AGA GTT TGA TCC TGG CTC AG	8-27	20	V1
13	27F	AGA GTT TGA TCC TGG CTC AG	8-27	20	V1
2	27F	AGA GTT TGA TCM TGG CTC AG	8-27	20	V1
3	27F	GA GTT TGA TCM TGG CTC AG	9-27	19	V1
4	357F	ACT CCT ACG GGA GGC AGC AG	338-357	20	V3
5	357F-1	CCT ACG GGR GGC AGC AG	341-357	17	V3
6	355F-2	ACWYCT ACG GRW GGC TGC	338-355	18	V3
7	355F-3	CA CCT ACG GGT GGC AGC	339-355	17	V3
9	357F	CCT ACG GGN GGC WGC AG	341-357	17	V3
10	357F	ACT CCT ACG GGA GGC AGC AG	338-357	20	V3
14	357F	CCT ACG GGA GGC AGC AG	341-357	17	V3
12	531F	GTG CCA GCM GCN GCG G	516-531	16	V4
16	534F	GTG CCA GCM GCC GCG GTA A	516-534	19	V4
8	937F	TTG ACG GGG GCC CGC AC	921-937	17	V6
15	984F	ACG CGA AGA ACC TTA C	969-984	16	V6

Note:

[a] Position of the primers in the 16S rDNA gene (based on *E. coli* position).

[b] The region in the 16S rDNA gene that the primers locate.

Some primers use ambiguous nucleotides such as the letter R or W. The IUPAC nucleotide code for the regular and ambiguous bases is presented in (**258**).

Similarly, reverse primers from N^0 1 to N^0 16 were organized according to the differences of their sequences, length and positions on the 16S rDNA gene (according to *E. coli* position) with the region from V3 to V9 (**Table 3.3**).

Table 3.3: Similarities and differences among the 16 reverse primers. Among all 16 reverse primers collected, primers N^0 1 & 10 are identical and primers N^0 2 & 13 are identical. The other primers are different in the ambiguous nucleotides, their position in the 16S rDNA gene, and their length.

N^0	Name	5'-->3' sequence	Position [a]	Length	Region Started [b]
1	515 R	TTA CCG CGG CTG CTG GCA C	515-533	19	V3
10	515 R	TTA CCG CGG CTG CTG GCA C	515-533	19	V3
2	518 R	A TTA CCG CGG CTG CTG G	518-534	17	V3
13	518 R	A TTA CCG CGG CTG CTG G	518-534	17	V3
3	518 R	W TTA CCG CGG CTG CTG G	518-534	17	V3
11	515 R	CG CGG CTG CTG G CAC	515-529	15	V3
4	785 R	TAC NVG GGT ATC TAA TCC	785-802	18	V4
9	785 R	GAC TAC HVG GGT ATC TAA TCC	785-805	21	V4
5,6,7	909 R	CCG TCA ATT YH T TTR AGT	909-926	18	V5
14	909 R	CCG TCA ATT CMT TTR AGT	909-926	18	V5
8	1492 R	TAC CTT GTT ACG ACT T	1492-1507	16	V9
15	1492 R	TAC GGY TAC CTT GTT AYG ACT T	1492-1513	22	V9
12	1100 R	GGG TTN CGN TCG TTR	1100-1114	15	V5
16	786 R	TAA TCT WTG GGV HCA TC AGG	786-806	20	V4

Note:

[a] Position of the primers in the 16S rDNA gene ((based on *E. coli* position)

[b] The region in the 16S rDNA gene that the primers cover.

3.1.1.2. Forward & Reverse primers tests:

We evaluated each Forward and Reverse primer collected above with two 16S rDNA databases: RDP (**259**), SILVA (**249**) (both were the versions of the year 2013) in order to see its coverage percentage in these databases (year 2013).

3.1.1.2.1. RDP and Silva primer testing programs:

a) RDP probe test (RDP 1.0 release):

In the RDP probe evaluation program, there are 4 parameters: Strain, Source, Size, and Quality (**Table 3.4**). In Strain and Source parameters, the Both Option was chosen to pick up Type & Non Type sequences and Uncultured & Isolates sequences. In the Quality parameter, Good Option was chosen because the Suspect Option indicated that the quality of the sequences in the probe search were not certain according to RDP website.

Then, two sets of parameters are generated:

- Set 1: using the Size Both Option → to cover all the sequences in the databases. There were 2,622,036 sequences in the RDP database.
- Set 2: using the Size ≥ 1200 bp Option → to cover the full-length sequences in the database. There were 1,258,845 sequences in the RDP full-length database.

Table 3.4: RDP Probe test parameters.

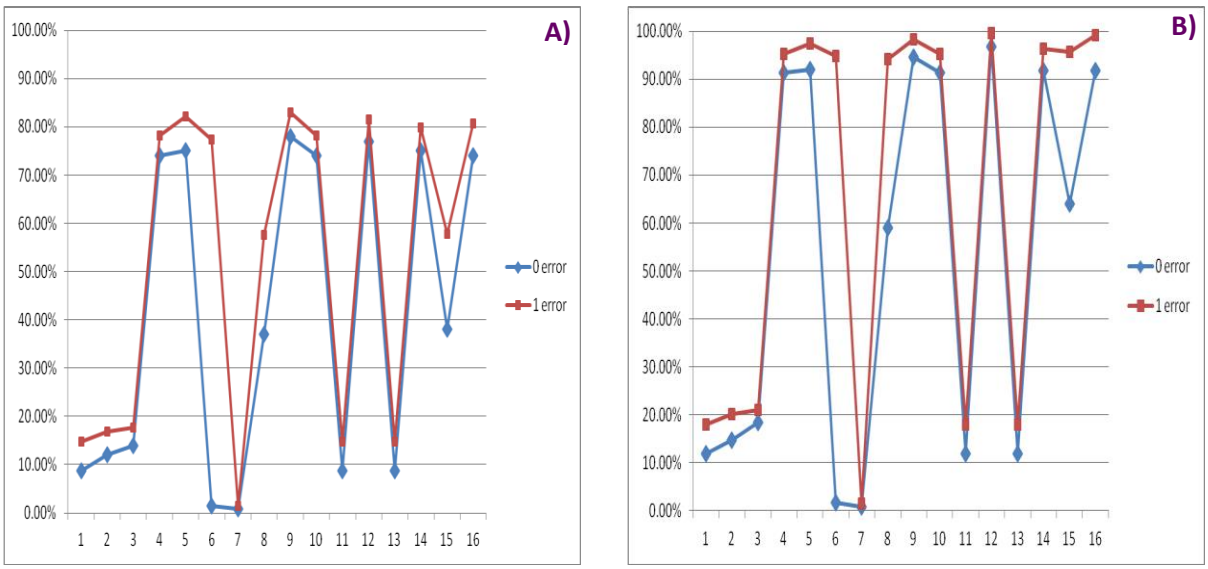
Strain	Type	Non Type	Both
Source	Uncultured	Isolates	Both
Size	≥1200 bp	<1200 bp	Both
Quality	Good	Suspect	Both

b) *SILVA Probe Test:*

In the SILVA Probe Test 3.0, the default database was high quality, non-redundant sequences, consisting of 281,797 sequences (much lower than the RDP database), called SILVA-r114 REFNR database.

3.1.1.2.2. Forward primer test results:

a) In RDP:



Total Database **Full-length Database**

Figure 3.1. The percentage matching score of 16 forward primers with RDP database.

Note:

A) All the sequences in database were chosen (Set 1 parameter).

B) Just full-length sequences (sequences that are ≥1200 bp) were chosen (Set 2 parameter).

To make this part easier to follow, forward primer N°1 will be called primer 1F, etc. Reverse primer N°1 will be called 1R, etc. Primer pairs N°1 will be called 1P, etc. (**Table 3.1**, **3.2** and **3.3**). Each forward primer was posted on the RDP Probe Testing Program Online (RDP 1.0 release) with the parameters described above. The searches were performed with 0 error and 1 error accepted in the primer. The number of 16S rDNA sequences for the Total Database is 2,622,036 and 1,258,845 for the Full-Length Database.

The results showed that for the Total Database (including sequences whose lengths are both <1200 bp and ≥1200 bp), the forward primer 4F (10F), 5F, 9F, 12F, 14F and 16F had high coverage (>70%) with 0 error and 1 error. Our lab primer, 3F had lower coverage than the others (about 14% and 17.7% with 0 error and 1 error, respectively). Particularly, primer 6F has very low coverage with 0 error allowed (1.5%) but raised to 77.3 % when 1 error is accepted (**Fig. 3.1** and **Tab. 3.5**).

For the Full-Length Database (including sequences whose lengths are ≥1200 bp), the primers 4F (10F), 5F, 9F, 12F, 14F and 16F had relatively high coverage (> 90%) with 0 and 1 errors. Our lab primer, primer 3F, had relatively lower coverage than others (about 18.3% and 20.9% with 0 and 1 errors, respectively). Once again, primer 6F has very low coverage when 0 error was accepted (1.6%) but raised to 94.8 % when 1 error was

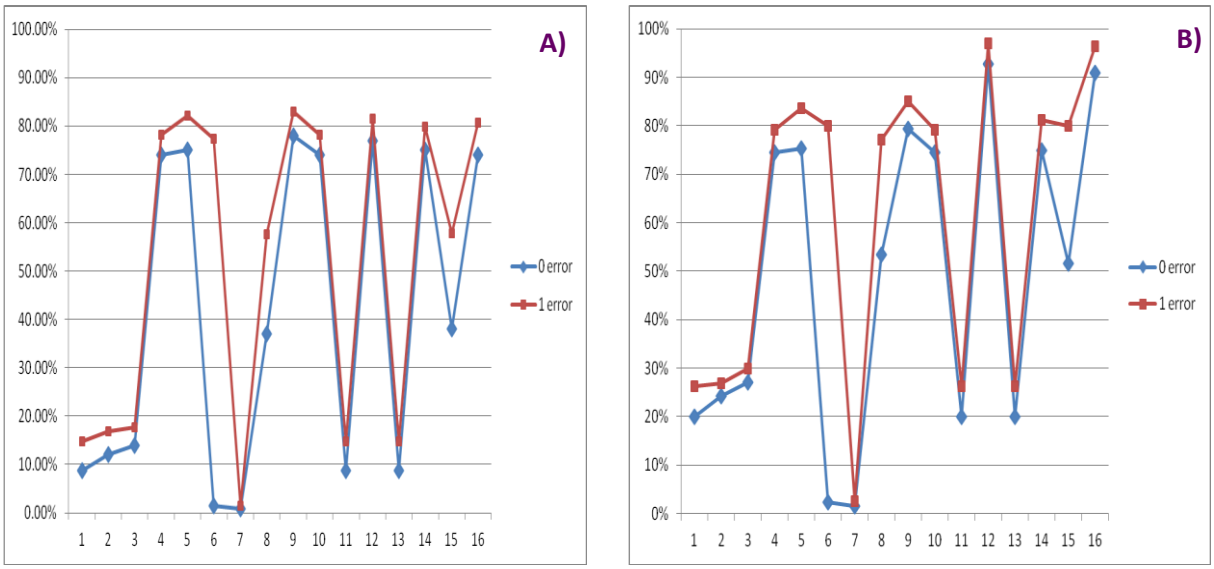
accepted. Primers 8F and 15F coverage improved approximately 30% from 0 error to 1 error.

In summary, for in-silico testing, the primers 4F (10F), 5F, 9F, 12F, 14F and 16F showed the highest coverage in both Databases with 0 and 1 error. Choosing these primers for 16S rDNA studies is likely to be recommended. The primers 6F, 8F, 15F should be used with caution due to their low coverage in the Databases and their aberrant behavior between 0 error and 1 error searching. The other primers, such as 1F, 2F, 3F, 7F, 11F and 13F should be considered carefully and further studied before use due to their low coverage in all the tested cases.

Table 3.5: Percentage coverage of 16 forward primers using the RDP Test Probe program.

N^{o}	**Total**		**Full length**	
	2,622,036		**1,258,845**	
	0 error	1 error	0 error	1 error
1F	8.8%	14.8%	11.9%	18.0%
2F	12.0%	16.9%	14.8%	20.1%
3F	14.0%	17.7%	18.3%	20.9%
4F	74.0%	78.1%	91.3%	95.3%
5F	75.0%	82.2%	92.0%	97.5%
6F	1.5%	77.3%	1.6%	94.8%
7F	0.9%	1.5%	0.8%	1.4%
8F	37.0%	57.7%	58.9%	94.1%
9F	78.0%	82.9%	94.5%	98.2%
10F	74.0%	78.1%	91.3%	95.3%
11F	8.8%	14.8%	11.9%	18.0%
12F	77.0%	81.6%	96.8%	99.5%
13F	8.8%	14.8%	11.9%	18.0%
14F	75.0%	79.8%	91.7%	96.3%
15F	38.0%	57.9%	63.9%	95.7%
16F	74.0%	80.7%	91.8%	99.2%

b) RDP versus SILVA:



RDP **SILVA**

Figure 3.2. Results the percentage matching score of 16 forward primers with RDP and SILVA database.

Note:
A) Result in RDP Probe test with all the sequences in database were chosen (Set 1 parameter).
B) Result in SILVA Probe Test.

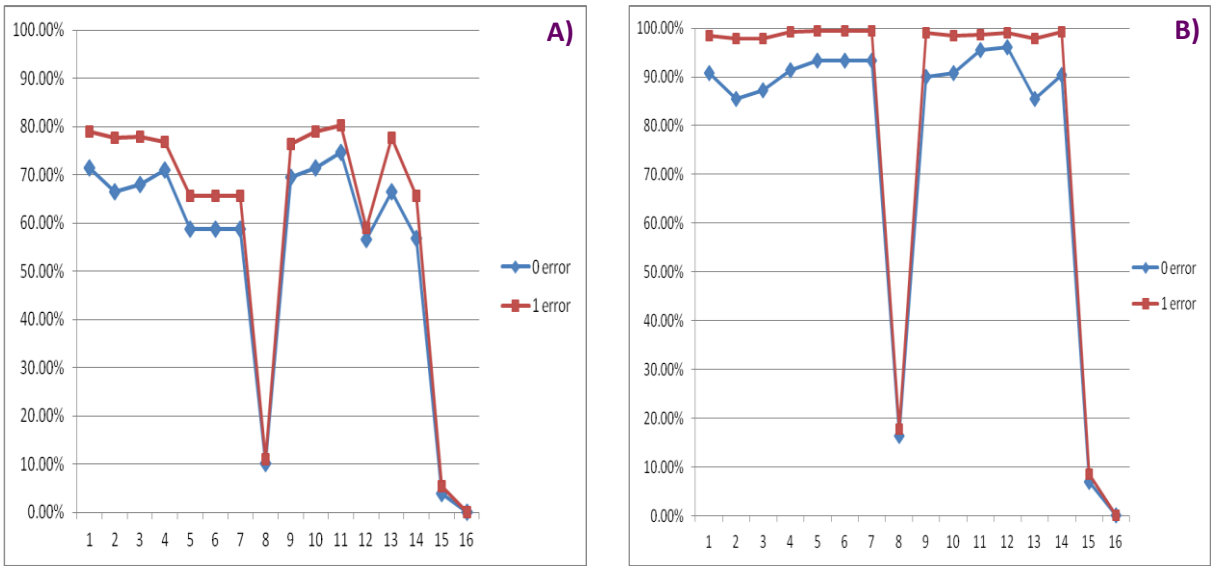
Primer test coverage in the SILVA Database-NR (non-redundant) was performed on the Silva website with 0 error and 1 error accepted in the primers. The results were similar for all the 16 forward primers (**Fig. 3.2**). The primers 4F (10F), 5F, 9F, 12F, 14F and 16F had the highest coverage in both RDP Total Database and SILVA Database (>70%). The lab primer 3F had lower coverage in the SILVA database ~ 30% with 0 error and 1 error. Again, the primers 4F (10F), 5F, 9F, 12F, 14F and 16F are recommended for 16S rDNA experiments (**Table 3.6**)

Table 3.6: Percentage coverage of 16 forward primers using RDP and SILVA Test Probe programs.

	RDP		SILVA-NR	
	2,622,036		281, 797	
N0	0 error	1 error	0 error	1 error
1F	8.8%	14.8%	20.0%	26.2%
2F	12.0%	16.9%	24.2%	26.9%
3F	14.0%	17.7%	27.1%	30.0%
4F	74.0%	78.1%	74.5%	79.2%
5F	75.0%	82.2%	75.4%	83.7%
6F	1.5%	77.3%	2.36%	80.0%
7F	0.9%	1.52%	1.57%	2.62%
8F	37.0%	57.7%	53.5%	77.2%
9F	78.0%	82.9%	79.3%	85.0%
10F	74.0%	78.1%	74.5%	79.2%
11F	8.8%	14.8%	2.0%	26.2%
12F	77.0%	81.6%	92.8%	97.0%
13F	8.8%	14.8%	20.0%	26.2%
14F	75.0%	79.8%	74.9%	81.1%
15F	38.0%	57.9%	51.5%	79.9%
16F	74.0%	80.7%	91.0%	96.4%

3.1.1.2.3. Reverse primer results:

a) In RDP:



Total Database **Full-length Database**

Figure 3.3. The percentage matching score of 16 reverse primers with the RDP database.
Note:
A) All the sequences in database were chosen (Set 1 parameter).
B) Just full-length sequences (sequences that are ≥1200 bp) were chosen (Set 2 parameter).

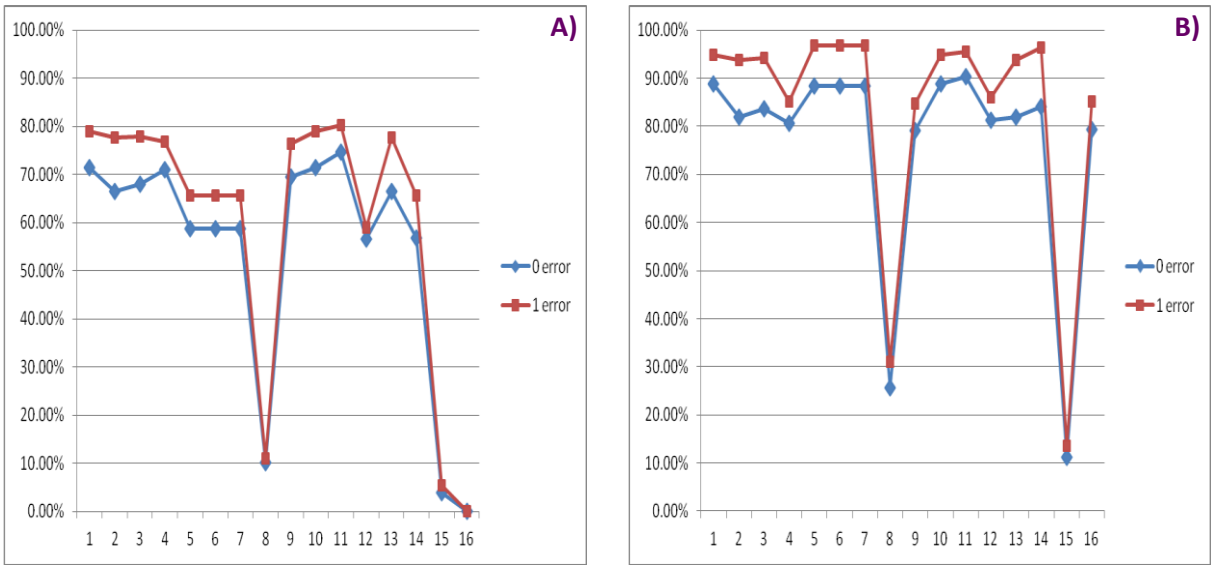
The same experiments were performed with the reverse primers with the RDP Probe test program. Using the Total Database, primers 1R, 4R, 10R, 11R had high coverage with both 0 error and 1 error searching (>70%). Our lab primer, 3R also has high coverage (68.1% and 77.8%). Primers 8R, 15R and 16R had very low coverage (<10%). Primers 2R, 3R, 5R and 13 coverages were improved (>70%) when 1 error was accepted in the search. For the Full-Length Database, all the primers had high coverage (> 80%) with 0 error and 1 error except primers 8R, 15R and 16R (**Fig. 3.3**, **Table 3.7**).

Overall, according to the results in RDP, the primers 1R (10R), 4R and 11R are highly recommended. The primers 8R, 15R and 16R should be carefully considered and studied more before use in due to their low coverage.

Table 3.7: Percentage coverage of 16 reverse primers using RDP Test Probe program and the RDP Database.

	Total		**Full length**	
	2,622,036		1,258,845	
N0	0 error	1 error	0 error	1 error
1R	71.4%	78.9%	90.8%	98.4%
2R	66.5%	77.6%	85.4%	97.8%
3R	68.1%	77.8%	87.2%	97.9%
4R	71%	76.9%	91.3%	99.2%
5R	58.8%	65.7%	93.4%	99.5%
6R	58.8%	65.7%	93.4%	99.5%
7R	58.8%	65.7%	93.4%	99.5%
8R	10.2%	11%	16.3%	17.7%
9R	69.5%	76.3%	90.1%	99.1%
10R	71.4%	78.9%	90.8%	98.4%
11R	74.6%	80.3%	95.6%	98.7%
12R	56.6%	58.9%	96.1%	99%
13R	66.5%	77.6%	85.4%	97.8%
14R	56.9%	65.6%	90.4%	99.3%
15R	4.02%	5.4%	6.93%	8.53%
16R	0%	0%	0%	0%

b) RDP versus SILVA:



RDP **SILVA**

Figure 3.4. The percentage matching score of 16 reverse primers with RDP and SILVA databases.

Note:
A) Result in RDP Probe test with all the sequences in database were chosen (Set 1 parameter).
B) Result in SILVA Probe Test.

The primer test with SILVA-NR Database (281,797 seqs) on the SILVA website yielded similar results to that on the RDP website (**Fig.3.4, Table 3.8**). The coverage of the primers improved to 20% in with 0 and 1 error searches in SILVA compared to the RDP Total Database. All the primers had high coverage (>80%) except primers 8R and 15R (<20%, **Table 3.8**). Primer 16R improved its coverage from 0% in RDP (both Databases) to 80% in SILVA.

Thus, for in-silico testing, the primers 1R (10R), 4R and 11R are recommended for 16S rDNA studies while the primers 8R and 15R should be further studied. For the case of primer 16R, the experiments were performed 3 times using different parameters but still yielded the same results.

Table 3.8: Percentage coverage of 16 reverse primers using RDP and SILVA Test Probe programs.

	RDP		SILVA-NR	
	2,622,036		281, 797	
N0	0 error	1 error	0 error	1 error
1R	71.4%	78.9%	88.9%	94.9%
2R	66.5%	77.6%	81.9%	93.8%
3R	68.1%	77.8%	83.7%	94.2%
4R	71.0%	76.9%	80.7%	85.2%
5R	58.8%	65.7%	88.4%	96.7%
6R	58.8%	65.7%	88.4%	96.7%
7R	58.8%	65.7%	88.4%	96.7%
8R	10.2%	11.0%	25.6%	31.1%
9R	69.5%	76.3%	79.1%	84.7%
10R	71.4%	78.9%	88.9%	94.9%
11R	74.6%	80.3%	90.3%	95.4%
12R	56.6%	58.9%	81.2%	86.1%
13R	66.5%	77.6%	81.9%	93.8%
14R	56.9%	65.6%	84%	96.3%
15R	4.0%	5.4%	11.2%	13.5%
16R	0.0%	0.0%	79.4%	85.2%

3.1.1.3. Primer pairs evaluation:

3.1.1.3.1. RDP and Silva primer pair testing programs:

The 16 primer pairs were evaluated using RDP Primer Test (version 2013) and SILVA Primer Pair Test (version 2013). In the RDP Primer Test, the parameters are as shown in Table 3.9 below.

Table 3.9: Parameters in RDP primer pairs test.

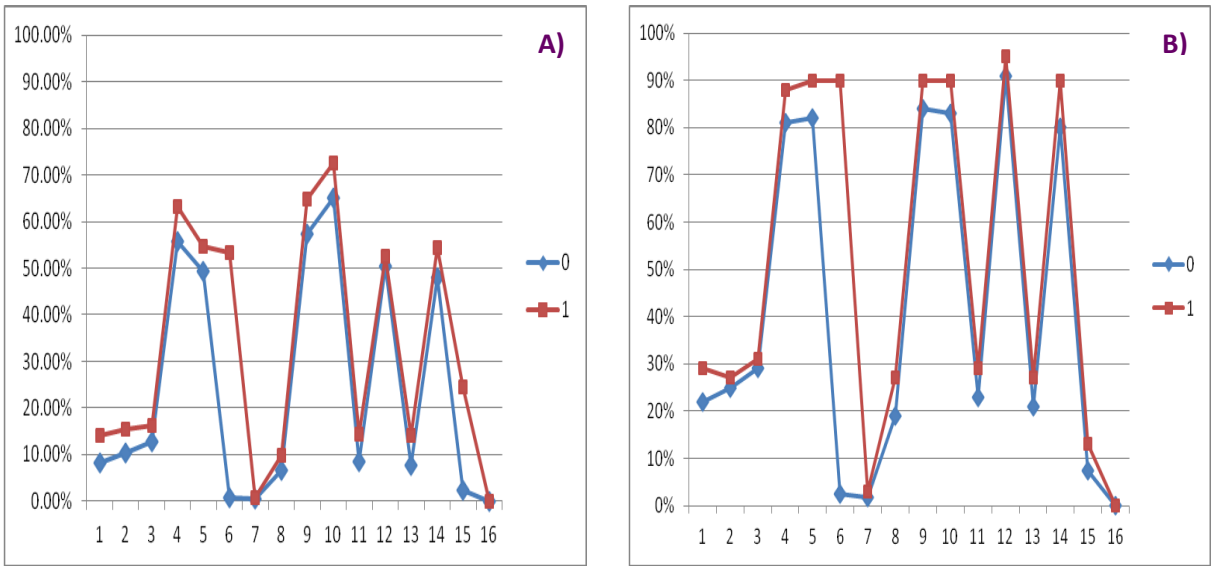
Parameters			
Strain	Type	Non Type	**Both**
Source	Uncultured	Isolates	**Both**
Size	≥1200	<1200	**Both**
Quality	**Good**	Suspect	Both

Note: Table 3.10 presents the parameters that were available on RDP Primer Test (version of the year 2012). The **bold** words are the parameters chosen in this analysis.

In the RDP Primer Test, each primer pair was tested twice with different error number allowed: 0 and 1 error.

In SILVA Primer Pair test, each primer pair was tested twice with different mismatch numbers allowed: 0 mismatch and 1 mismatch allowed (the mismatch is set to occur not in the 5-last nucleotides at the 3' end of the primer. This is because if there is any mismatch at or near the 3' end of the primer, the chance for a successful PCR reaction to happen would be low) (SILVA Primer Pair test recommendation).

3.1.1.3.2. Results:



RDP (2,622,036 seqs) **SILVA (281, 797 seqs)**

Figure 3.5. The percentage matching score of 16 primer pairs with RDP and SILVA databases.

Note:

A) Result in RDP Probe test with all the sequences in database were chosen (Set 1 parameter).

B) Result in SILVA Probe Test.

The same patterns were observed between the two database (**Fig. 3.5**) and the two primer pairs testing programs using two databases gave similar results of the coverage for each primer pair.

In the RDP database**,** primer pairs 4P, 5P, 9P, 10P, 12P and 14P had medium coverage (>60%) as observed with SILVA (>80%) (**Table 3.14**). Primer pair 6P had low coverage in RDP with 0 error (2.5%) and in SILVA with 0 mismatch (0.87%), but increased dramatically with 1 error allowed in RDP (53.2%) and 1 mismatch in SILVA (90%). This is probably caused by the use of the forward primer 6F in Forward Primer test section.

We chose the forward, reverse primers and primer pairs that had high coverage with both 0 & 1 error to continue the analysis. They are 4F, 5F, 9F, 12F, 14F and 16F for forward primers; 1R, 2R, 3R, 4R, 5R, 9R and 11R for reverse primers and 4P, 5P, 9P, 10P, 12P and 14P. (Note: among the forward primers, 4F and 10F are the same and for the reverse primers 1R=10R, 13R=2R).

3.1.1.4. Bacteria, Archaea & Eukarya evaluation:

To survey if the 16S universal primers have the potential to amplify organisms such as Archaea and Eukarya, we examined the primers chosen above. The tests were performed on RDP and SILVA website, following their guide to choose the required database (in this case Bacteria, Archaea and Eukarya were deposited into different databases).

3.1.1.4.1. Forward primers:

For the Forward primer, using the RDP Primer Test, primer 4F and 14F had high coverage (> 70%) in Bacteria but very low coverage in Archaea in both 0 and 1 error searches (≈ 0.1%) (**Table 3.10**). Primer 5F and 9F had high coverage in Bacteria (>70%) with 0 error and 1 error search, very low coverage in Archaea with 0 error search (0.3%) and high coverage in Archaea with 1 error search (≈ 46.6%). Primers 12F and 16F had high coverage in Archaea with 0 and 1 error search (≈ 47% and 83%, respectively).

In the SILVA Probe Test program, the results are similar for each primer for Bacteria and Archaea. For Eukarya, the primers had high coverage (> 80%) (**Table 3.11**). Using silico testing, primers 4F&14F appear to be a good choice for 16S rDNA study while primers 5F&9F should be further studied and primers 12F&16F should be avoided.

Table 3.10: Percentage of Bacteria, Archaea, Eukarya domain of chosen Forward primers in the RDP database.

N⁰	**Bacteria**		**Archaea**	
	2,493,318		128,216	
	0 error	1 error	0 error	1 error
4F	77.8%	82.1%	0.03%	0.03%
5F	79.1%	84.0%	0.3%	46.6%
6F	1.5%	81.3%	0.0%	0.03%
9F	81.6%	84.7%	0.3%	46.8%
12F	78.8%	81.5%	47.6%	83.4%
14F	78.7%	83.9%	0.1%	0.9%
16F	35.2%	80.6%	47.2%	83.0%

Table 3.11: Percentage of Bacteria, Archaea, Eukarya domain of chosen Forward primers in the SILVA database.

Nº	Bacteria		Archaea		Eukarya	
	239,346		10,851		31,600	
	0 error	1 error	0 error	1 error	0 error	1 error
4F	88.0%	93.0%	0.0%	0.0%	0.0%	0.0%
5F	89.0%	96.0%	0.3%	64.0%	0.06%	0.2%
6F	2.8%	94.0%	0.0%	0.0%	0.0%	2.2%
9F	93.0%	97.0%	0.3%	64.0%	0.06%	0.3%
12F	95.0%	97.0%	55.0%	94.0%	90.0%	94.0%
14F	88.0%	95.0%	0.07%	1.7%	0.05%	0.1%
16F	93.0%	94.0%	54.0%	97.0%	88.0%	93.0%

3.1.1.4.2. Reverse primers:

For the reverse primers, in the RDP Primer Test:

- Primer 2R&3R had high coverage (≥ 70%) in Bacteria with 0 and 1 error searches but very low coverage in Archaea (≈ 0.3%) in 0 error search. However, with 1 error allowed, the coverage of these primers increased up to ≈ 25% (**Table 3.12**).
- Primers 1R and 11R had high coverage in Bacteria (> 75%) with 0 and1 error searches, very low coverage in Archaea with 0 error search (0.1%) and medium coverage with 1 error (≈ 47%).
- Primers 5R&14R had medium coverage in Bacteria (≥ 60%) with 0 and 1 error searches, very low coverage in Archaea with 0 error (0.1%) and high coverage in Archaea with 1 error (≈ 70%).
- Primer 12R had medium coverage in Bacteria (≈ 60%), very low coverage in Archaea with 0 errors (0.02%) and increased up to ≈ 20% with 1 error.
- Primers 4R&9R had high coverage in Bacteria (> 70%) with 0 and 1 error search, and high coverage in Archaea (> 80%) with 0 and 1 error.

In the SILVA Probe Test program (**Table 3.13**), in Bactera & Archaea coverage, the results are similar for each primer with:

- Primers 2R and 3R had high coverage in Bacteria (> 80%) 0 and 1 error searches, very low coverage in Archaea (> 0.1%) with 0 error and increased up to ≥ 30% in 1 search.
- Primers 1R and 11R had high coverage in Bacteria (≥ 90%) 0 and1 error, very low coverage in Archaea (> 1.5%) with 0 error and medium coverage with 1 error (55%).

- Primers 5R and 14R had high coverage in Bacteria (≥ 90%) with 0 and 1 error searches, very low coverage in Archaea with 0 error (< 1%) and very high coverage in Archaea with 1 error (> 85%).
- Primer 12R had high coverage in Bacteria (> 95%) with 0 and 1 error searches, very low coverage in Archaea with 0 error (0.08%) and medium coverage ≈ 50% with 1 error.
- Primers 4R and 9R had both high coverage in Bacteria (> 90%) and high in Archaea (≥ 90%) with 0 and 1 error searches.
- For Eukarya, the primers 1R, 2R, 3R, 5R, 11R and 14R had high coverage (≥ 80%) and primers 4R, 9R, 12R and 16R had very low coverage (<5%).

In conclusion:

For each primer, the RDP and SILVA results are in agreement for Bacteria and Archaea with 0 and 1 error, although coverage percentage of all the primers is slightly higher using SILVA.

Using in-silico testing, primers 2R&3R appeared to be a good choice for 16S rDNA studies due to their high coverage in Bacteria and lowest coverage in Archaea among all the primers. Its high coverage in Eukarya, however can be problematic when analyzing the bacterial community. However, note that the database of Eukarya in SILVA is very small (31 seqs). Primer 12R is also a good choice with its medium coverage in Bacteria and very low coverage in Archaea and Eukarya. Primers 4R and 9R should be avoided due to their high coverage in Archaea.

Table 3.12: Percentage of Bacteria, Archaea of chosen Reverse primers in the RDP database**.**

N⁰	**Bacteria**		**Archaea**	
	2,493,318		128,216	
	0 error	1 error	0 error	1 error
1R	75.1%	83.0%	0.93%	47.4%
2R	69.9%	80.3%	0.35%	23.9%
3R	71.6%	80.5%	0.37%	25.9%
4R	70.3%	76.4%	83.8%	87.7%
5R	61.8%	65.5%	0.71%	69.7%
9R	68.8%	75.7%	82.8%	87.4%
11R	78.4%	82.0%	0.95%	47.9%
12R	59.5%	61.0%	0.02%	18.7%
14R	59.8%	65.4%	0.71%	68.8%

Table 3.13: Percentage of Bacteria, Archaea of chosen Reverse primers in the SILVA database.

N⁰	Bacteria		Archaea		Eukarya	
	239,346		10,851		31,6	
	0 error	1 error	0 error	1 error	0 error	1 error
1R	93 %	97%	1.5%	55%	88%	93%
2R	85%	97%	0.42%	32%	82%	94%
3R	87%	97%	0.44%	37%	86%	94%
4R	91%	96%	90%	96%	1.1%	2%
5R	92%	97%	0.95%	87%	93%	97%
9R	89%	95%	88%	95%	0.03%	0.84%
11R	94%	97%	1.5%	55%	89%	94%
12R	96%	99%	0.08%	46%	0.02%	5%
14R	89%	97%	0.92%	86%	77%	96%
16R	90%	96%	88%	96%	0%	0.42%

3.1.1.4.3. Primer pairs:

Table 3.14 presents the percentage coverage of Bacteria, Archaea, Eukarya domain of 16 primer pairs in 2 databases: RDP and SILVA.

a) In RDP:

- Primer pairs 4P, 10P and 14P had medium and high coverage in Bacteria with 0 and 1 error (from 48.0-72.6%), and very low coverage in Archaea (< 0.1%) with 0 and 1 error.
- Primer pairs 5P, 9P and 12P had medium coverage in Bacteria (from 49.4-64.7%), very low coverage in Archaea (< 0.1%) with 0 error and increasing from 13.8 to 41.2% with 1 error.
- Primer pairs 1P, 2P, 3P, 8P, 11P and 13P had low coverage in Bacteria (from 7.6 to 16.1%) with 0 and 1 error, very low coverage in Archaea (< 0.03%) with 0 and 1 error.
- Primer pair 6P had very low coverage in Bacteria with 0 error, but significantly increased with 1 error accepted (from 0.87 to 53.2%), very low coverage of Archaea (< 0.03%) with 0 and 1 error.
- Primer pair 16P had no coverage (0.0%) of Bacteria, Archaea and Eukarya.

b) In SILVA:

- Primer pairs 4P, 10P and 14P had high coverage in Bacteria with 0 and 1 error (≥ 80%), and very low coverage in Archaea (≤ 0.12%) with 0 and 1 error. These primer pairs have very low coverage in Eukarya (0.0-0.12%) with 0 and 1 error.

- Primer pairs 5P, 9P and 12P had high coverage in Bacteria (≥ 80%), very low coverage in Archaea (≤ 2.2%) with 0 and 1 error, except 9P has a medium coverage of Archaea (61%) with 1 error. These primer pairs have very low coverage in Eukarya (0.0-0.13%) with 0 and 1 error.
- Primer pairs 1P, 2P, 3P, 8P, 11P and 13P has low coverage in Bacteria (from 19-31%) with 0 and 1 error, little coverage in Archaea (0.0%) with 0 and 1 error. These primer pairs have little coverage in Eukarya (0.0-0.0%) with 0 and 1 error.
- Primer pair 6P has very low coverage in Bacteria with 0 error, but significantly increased with 1 error accepted (from 2.5 to 90%), and no coverage of Archaea (0.0%) with 0 and 1 error. This primer has a low coverage on Eukarya but was the highest among the primers (2.2%) with 1 error.
- Primer pair 16P has no coverage (0.0 %) of Bacteria, Archaea and Eukarya.

In conclusion:

Results in both RDP and SILVA were in agreement among all primer pairs for Bacteria and Archaea, with coverage in SILVA greater than that in RDP, probably due to the greater number of sequences of RDP. RDP did not have a significant database for Eukarya at that time so there were no results of Eukarya using RDP.

Table 3.14: Percentage coverage of Bacteria, Archaea, Eukarya domain of 16 primer pairs in 2 databases.

Nº	Domain	RDP		SILVA	
		0	1	0	1
1	B	8.20%	14.12%	22.00%	29.00%
	Ar	0.00%	0.01%	0.00%	0.00%
	Eu			0.00%	0.00%
2	B	10.40%	15.52%	25.00%	27.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%
3	B	12.70%	16.16%	29.00%	31.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%
4	B	55.63%	63.13%	81.00%	88.00%
	Ar	0.03%	0.03%	0.00%	0.00%
	Eu			0.00%	0.00%
5	B	49.44%	54.62%	82.00%	90.00%
	Ar	0.03%	30.53%	0.00%	0.74%
	Eu			0.00%	0.13%
6	B	0.87%	53.2%	2.50%	90.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	2.20%
7	B	0.49%	0.89%	1.70%	2.90%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.06%
8	B	6.70%	10.00%	19.00%	27.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%
9	B	57.20%	64.70%	84.00%	90.00%
	Ar	0.32%	41.20%	0.29%	61.00%
	Eu			0.00%	0.08%
10	B	65.00%	72.60%	83.00%	89.00%
	Ar	0.00%	0.03%	0.00%	0.00%
	Eu			0.00%	0.00%
11	B	8.40%	14.40%	23.00%	29.00%
	Ar	0.00%	0.01%	0.00%	0.00%
	Eu			0.00%	0.00%
12	B	50.30%	52.50%	91.00%	95.00%
	Ar	0.01%	13.80%	0.06%	2.20%
	Eu			0.01%	0.01%
13	B	7.60%	14.00%	21.00%	27.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%
14	B	48.00%	54.50%	80.00%	90.00%
	Ar	0.03%	0.60%	0.00%	0.12%
	Eu			0.05%	0.06%
15	B	2.40%	4.80%	7.30%	13.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%
16	B	0.00%	0.00%	0.00%	0.00%
	Ar	0.00%	0.00%	0.00%	0.00%
	Eu			0.00%	0.00%

3.1.1.5. Discussion:

16S rDNA primers choice can affect downstream analyses due to (i) their different coverage in database, (ii) accuracy of assigned classification and diversity metric based on the amplicon generated by the primers and (iv) the chance of amplifying Archaea and Eukarya.

Primer pair 4P, which covers the region V3-V4 is the most promising primer pair due to its high accurate classification in previous studies (**335, 336, 337**). Carlos et al. (2010) showed that 347F/803R was the most suitable pair of primers for classification of foregut 16S rRNA genes and suitable for analyses of other complex microbiomes due to its high coverage in RDP database (from 84.9% to 98.7% with 0 and 1 mismatches) (**335**). In another study, the primer set of V3F and V4R covering regions V3V4 showed the best performance with i) less bias between S-V3V4 and S-V4V3 at both phylum/class and genus levels; ii) V3F and V4R covered more genera than other primer sets (**336**). Moreover, primer pair (S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21) yielding a 464 bp amplicon of the V3-V4 region should be the most promising bacterial primer pair among 175 primers and 512 primer pairs with *in silico* evaluated coverage and phylum spectrum with respect to the SILVA 16S/18S rDNA non-redundant reference dataset (SSURef 108 NR) and experimentally comparing with the 16S rDNA fragments from directly sequenced metagenomes (**337**).

In silico, testing of S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 failed to detect seven bacterial phyla (Hyd24-12, GOUTA4, *Armatimonadetes*, *Chloroflexi*, BHI80-139 and Candidate divisions OP11 and WS6). If one mismatch is tolerated (A: 64.6%, B: 94.5%, E: 0.1%), amplification of four additional phyla is likely (*Chloroflexi*, BHI80-139, Hyd24-12 and GOUTA4). In the *in silico* evaluation, S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 failed to detect SAR 11 clade 1. However, experimental evaluation showed that the primer pair is able to amplify this taxonomic group.

This can be explained by the increased coverage of up to 97% if one mismatch is allowed. A closer look at the primer target position of the reverse primer reveals an internal mismatch position towards the 5′ end. The results demonstrate that S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 provides a representation of the bacterial diversity down to genus levels and illustrates that an internal mismatch towards the 5′ end can be tolerated by standard PCR (**337**). The study indicated that primer re-evaluation coverage, including forward, reverse and primer pairs, on up-to-date 16S rDNA databases should be performed

before an experiment to examine which taxa can be failed to detect compared to experimental results.

Comparisons of species richness estimates of simulated pyrosequencing-generated fragments in 16S rDNA indicated that fragments encompassing the V4, V5+V6, and V6+V7 regions (generated using primer pairs 530F-805R, 805F-1046R, and 967F-1220R) provided estimates comparable to those obtained with the nearly full-length fragment (**338**).

In contrast, other amplicon generating other variable regions such as V1-V2, V1-V3, V5-V6, V7-V8 should be in caution when using for 16S rDNA –based study. It was showed that fragments encompassing the V1 and V2 (V1+V2) region and the V6 region (generated using primer pairs 8F-338R and 967F-1046R) overestimated species richness; fragments encompassing the V3, V7, and V7+V8 hypervariable regions (generated using primer pairs 338F-530R, 1046F-1220R, and 1046F-1392R) underestimated species richness; and fragments encompassing the V4, V5+V6, and V6+V7 regions (generated using primer pairs 530F-805R, 805F-1046R, and 967F-1220R) provided estimates comparable to those obtained with the nearly full-length fragment (**338**).

Several genera were detected more abundant than others according to different hypervariable regions studied. For example, the genera Prevotella, Fusobacterium, Streptococcus, Granulicatella, Bacteroides, Porphyromonas and Treponema were abundant when the V1-V3 region was targeted, while Streptococcus, Treponema, Prevotella, Eubacterium, Porphyromonas, Campylobacter and Enterococcus predominated in the community generated by V4-V6 primers, and the most numerous genera in the V7-V9 community were Veillonella, Streptococcus, Eubacterium, Enterococcus, Treponema, Catonella and Selenomonas. Targeting the V4-V6 region failed to detect the genus Fusobacterium, while the taxa Selenomonas, TM7 and Mycoplasma were not detected by the V7-V9 primer pairs **(339**).

Oppositely, the community fingerprints generated by V1-V3 and V7-V9 primers provided results similar to Sanger sequencing and were recommended by the author for using in 16S rDNA-based study (**339**). In general, when studying complex microbial communities in particular environment, two or more different amplicons should be examined separately and compared to each other to avoid the bias in classification and diversity metric analyses causing by single short amplicon.

3.1.2. Initial pipeline test for raw sequencing data:

Three pyrosequencing runs were obtained on a GS Junior (described in **Materials & Methods**). The signal processing of each run is filtered in two ways: **Amplicon filter** and **Shotgun filter.** Sequences that have these characteristics are considered as errors in pyrosequencing process (**241**):

1) Sequences that have length ≤200 and ≥600.

2) Sequences that have homopolymer >=8 bp.

3) Sequences that have more than 1 ambiguous base (N base)

4) Sequences that have more than 1 mismatch in the barcodes with 15 barcodes (or MIDs or Tags) were used in the 3 runs.

5) Sequences that have more than 1 mismatch to the forward primer (518R).

Base on these criteria, the 3 run sequences were analyzed for their errors using two bioinformatic programs: Qiime and Mothur (**240**, **262**) (**Table 3.15**). Based on the results, we can see that the Shotgun filter data contain more errors than Amplicon filter data in the 3 runs. Hence, **Amplicon data** were chosen for further analyses.

Table 3.15: Characteristics of the three GS Junior sequencing runs in Amplicon filter and Shotgun filter procedures.

3 runs	Total sequences	1.Length		2.Homopolymer	3.Ambiguous base (N)	4.Barcode mismatch	5.Primer mismatch
		≤200	≥600	≥8	≥1	≥1	≥1
1st Shotgun	166,010	338	9,039	442	34,004	3,871	1,358
1st Amplicon	136,776	110	181	50	16,255	2,533	954
2nd Shotgun	135,368	290	6,337	241	32,184	8,35	1,24
2nd Amplicon	102,826	112	95	19	16,484	5,067	700
3rd Shotgun	165,764	164	10,158	623	35,702	7,417	1,559
3rd Amplicon	118,790	42	160	37	14,456	4,212	965

Note:

- The **3 runs** column is the runs obtained from GS Junior sequencing with Shotgun or Amplicon filters. For example, 1st Shotgun is the first run with Shotgun filter.
- The **Total sequences** column is the sequences obtained from each run with Shotgun or Amplicon filters.
- The **Length** column is N^0 of sequences that have length ≤200 and ≥600.
- The **Homopolymer** column is N^0 of sequences that have length >=8 bp.
- The **Ambiguous base (N)** column is N^0 of sequences that have N base call
- The **Barcode** column is N^0 of sequences that have more than 1 mismatch to the barcodes.
- The **Primer mismatch** column is N^0 number of sequences that have more than 1 mismatch to the forward primer (518R).

3.1.2.1. Pipelines of 454 data initial processing:

Pipelines for 454 data initial processing software were collected from several studies, including Greengene, RDP, Galaxy, Qiime, Mothur, Pyrotagger, PANGEA, CloVR-16S, Prinseq, VAMP, WebMGA, MEGAN (**240, 243, 259, 347-354**). Among those, RDP, Galaxy, Qiime and Mothur were chosen to test because they are frequently used in publications for their ability to filter data by length, homompolymers, to split barcodes (Tags or Mids), to match forward primer, and to trim by quality score.

To test the performance of the chosen pipelines (**Fig. 3.6**), one raw sequence fasta file was put into the four pipelines, the output number of sequences were found to be different (data not shown).

These four pipelines trim sequences by (i) length, (ii) homopolymer, (iii) ambiguous base, (iv) splitting barcodes and (v) forward primers with mismatch allowed (**241**). However, the orders of these trimming steps are not the same in 4 pipelines with can affect the output sequences. For example, in RDP initial trimming process, there is no homopolymer trimming step (**Fig. 3.7**).

After setting 3 pipelines with the same parameters (**Fig. 3.7**), the output numbers of sequences were different (data not shown):

- Length, homopolymer, ambiguous base filtering are identical for the 4 pipelines.
- Split Mid and match forward primer steps: **RDP, Qiime**, and **Mothur** were chosen for further test.

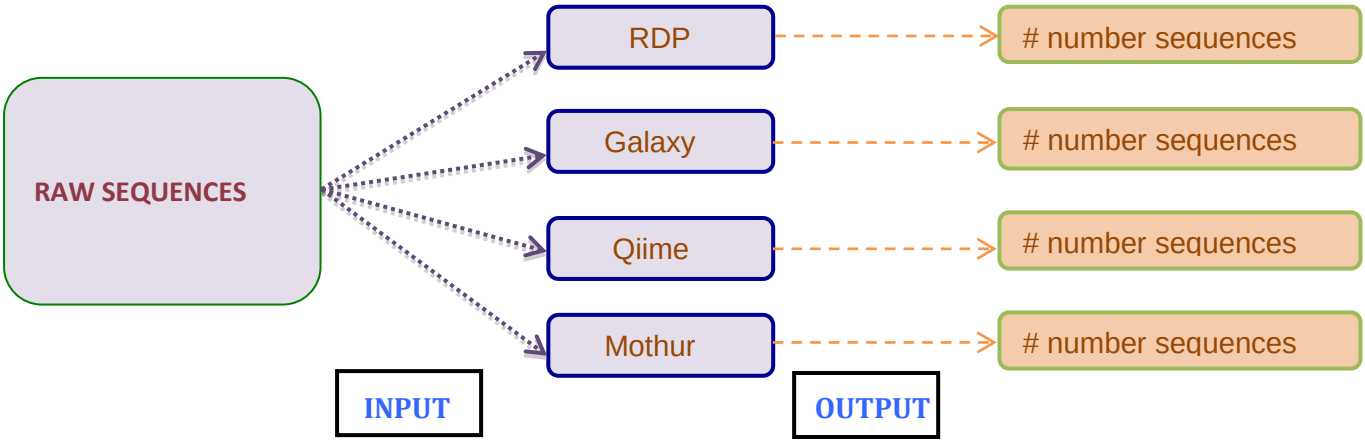


Figure 3.6. Testing the performances of four chosen pipelines.

The differences of the output number of sequences were found to be due to the order of sequence filtering steps (**Fig 3.7**).

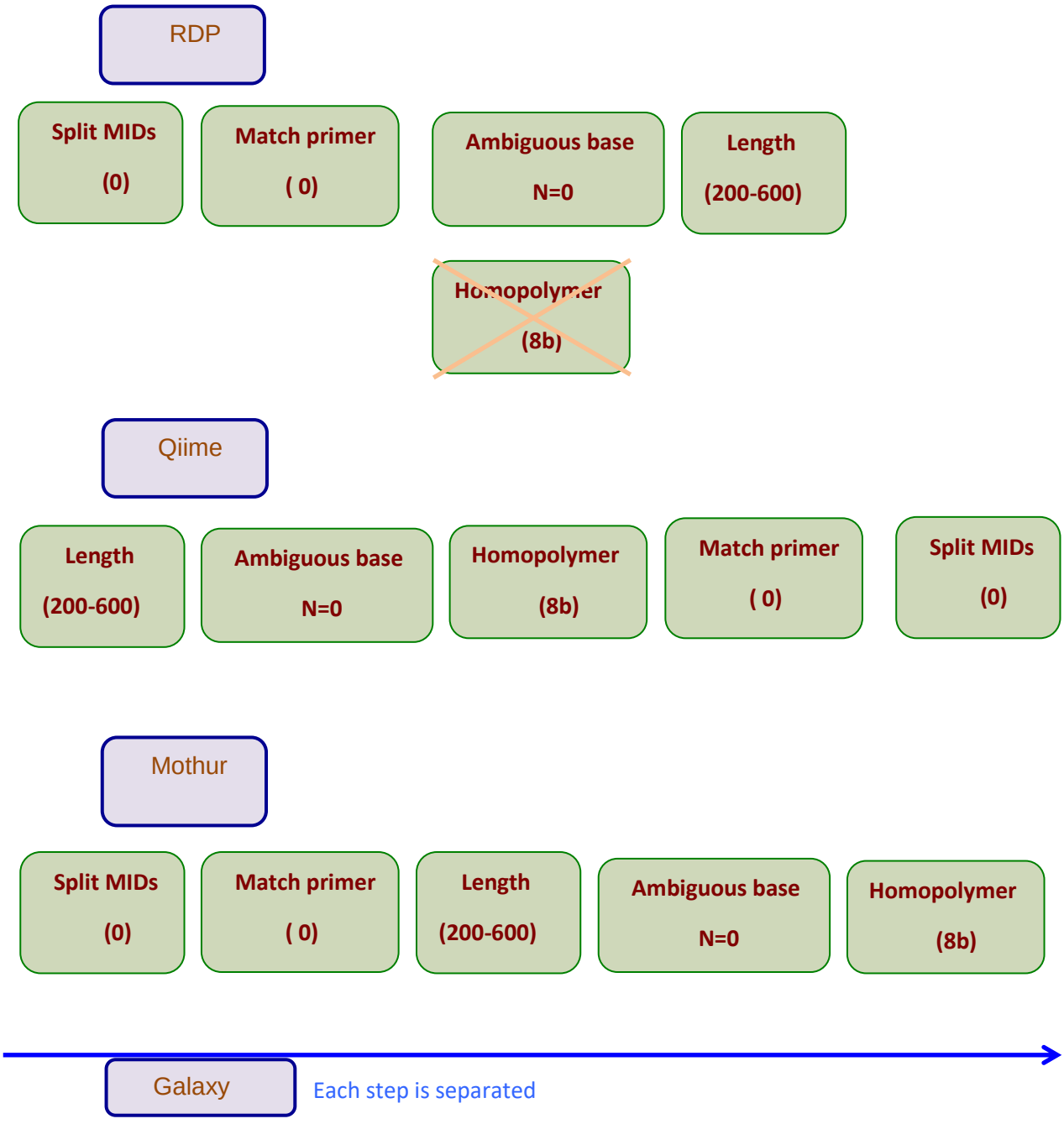


Figure 3.7. The different orders of sequences filtering steps of 4 pipelines.

Note: The blue arrow is the orders of the filtering steps. For example, in RDP, sequences were split by the Barcodes (or Mids) first, then they are filtered by the forward primer, ambiguous base and length, etc. the number is the setting for each pipeline. All the filtering steps were set exactly the same among these four pipelines with choosing sequences that have length from 200 to 600 nt, splitting Mids with 0 error, matching primer with 0 error, allowing no ambiguous base (N=0) and homopolymer ≥8. Among 4 pipelines, Galaxy allows to perform each filtering step separately while the others do not.

Mid (Barcode or Tag) is 10-nucleotide sequence presents at the beginning of each read. They are used to separate different samples in one run. There are 15 Mids used in this study. By allowing 1 and 2 errors in Mids and forward primers, 3 programs produced different numbers of sequences as they use different algorithms (**Fig. 3.8**).

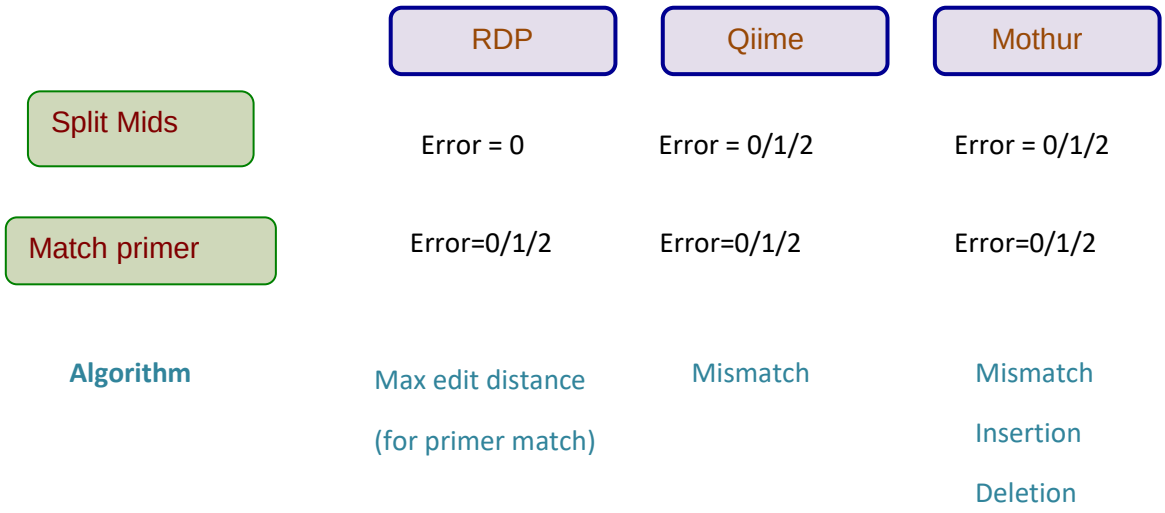


Figure 3.8. Algorithms of RDP, Qiime and Mothur in spliting Mids and matching forward primers. Note: RDP do not allow any error in barcode.

With Split Mids =0 & Primer mismatch =0, the number sequences from Mothur and Qiime are identical (data not shown), but RDP give different result.

Allowing 0 errors in Barcodes and 0 errors in primer match were chosen in the initial data processing by Qiime and Mothur. The errors (insertions, deletions or mismatches) that exist in Barcodes or Mids and in the forward primer reflect the error of light signals at the beginning of the sequences. Allowing errors in Barcodes and forward primers will yield more sequences (~1000-2000 / 136,776 sequences, based on tested data) but the data would contain a bad base at the beginning of the sequences and there could affect the downstream data analysis. Zero errors in barcode and in the forward primers were chosen.

3.1.2.2. Discussion:

Microbiologists who work with high-throughput data sequencing often treat raw data on available pipelines. Initial raw data treatment appears to be simple and easy to perform. However, each filtering step should be carefully considered before being performed to avoid the errors of pyrosequencing that can affect downstream analyses such as OTU grouping (**230**).

3.1.3. Quality score trimming program tests: Mothur, Galaxy and Condetri.

3.1.3.1. Parameters of each program:

454 data can have bad quality scores at the 3' end of the sequences, so that sequences should be trimmed from the 3' end. Since Qiime and RDP do not have 3' end Quality Score trimming, three other programs were chosen. They are Mothur, Galaxy and Condetri (**240**, **243**, **263**). Condetri trims sequences from the 3' end (**243**).

The **Table 3.16** describes the Trimming algorithm of these 3 pipelines with windows size, step size and quality score averages. The Input Data used to test different parameters of three chosen pipelines were the *E. coli* sequences of DNA extracted from strain MG1655, then followed by the same procedure of the PCR conditions for the sediment samples (**Materials** & **Methods**), and then mixed with PCR products of the sediment DNA samples in the 1st run of pyrosequencing with MID 15. Raw sequences of the 1st run were trimmed with 0 errors in barcodes (MIDs) and forward primer, excluding sequences that have length ≤ 200 and ≥ 600, homopolymer $\leq$ 8b, amplicon data type as describe in **Table 3.17** performed by Mothur. After trimming, two files types were produced: fasta and qual files.

Table 3.16: Different window size of 3 pipelines: Galaxy, Mothur and Condentri.

	Sliding widow			
	Direction	Window size (bases)	Step size (base)	Quality score average
Galaxy	3’	5	1	25
		10	1	25
		25	1	25
Mothur	3’ (?)	5	1	25
		25	1	25
		50	1	25
		a 50	1	35
Condetri	3’	**b** 5	1	25
		c 5	1	25

Note:

a- Mothur with Window size 50 and Quality average 35 are often used in publications (**241**).

b- Condetri (parameter f0.8): sequences after 3’ trimming are filtered with 80% of bases having Quality score ≥ 25.

c- Condetri (parameter f0.5): sequences after 3’ trimming are filtered with 50% of bases having Quality score ≥ 25.

After 3’ end trimming, a good program should produce the sequences satisfying at leatst 3 criteria:

i- Sequences length does not vary much.

ii- Average quality score per base should be greater than 10.

iii- The sequences should contain 90% of bases that have Quality Scores ≥ 25.

Table 3.17: Input file data for testing different parameters of Galaxy, Mothur and f parameters of Condetri.

Input Data (fasta&qual)	MID (15^{th})	Forward Primer	Length	Homopolymer	Data Filter type	Sample name
	0 error	0 error	200-600	≤ 8b	Amplicon	Control (*E.coli*)

The input files for quality trimming were 1_amplicon.raw.trim.control.fasta& 1_amplicon.raw.trim.control.qual. The number of sequences was 3487.

3.1.3.2. Trimming:

Different parameters of quality score trimming from Galaxy, Mothur and Condetri were performed with the Input file. The parameter principle is mainly based on sliding window and average quality score calculated in the sliding window. For example, Q25_s5 means a sliding window of 5 nt was created throughout each read and the quality score averages were calculated. If quality score averages in a 5nt window is > 25, the sliding window will continue throughout the sequence. If quality score averages in a 5nt window are < 25, the sliding window is terminated and this part of the sequence will be trimmed off. After going through different parameters, sequences were checked again by Quality cutoff 90% performed by Galaxy for the number of sequences that remained and the percentage of bad sequences were removed after different quality score trimming parameters are presented in Table 3.18.

Table 3.18: Different Sliding Window trimming (s) parameters and Quality cutoff 90%.

Pipeline	Parameters	Number of output sequences	Quality cutoff _90 [a]			
			Q25		Q20	
Galaxy	Q25_s5	3487	2656	23%	3266	6%
	Q25_s10	3487	2648	24%	3285	5%
	Q25_s25	3487	2641	24%	3327	4%
Mothur	Q25_s5	3485	3407	2%	3479	0%
	Q25_s25	3486	3103	10%	3388	2%
	Q25_s50	3485	2685	22%	3390	2%
	Q35_s50	3397	3244	4%	3386	0%
Condetri	Q25_s5_f0.8	3251	2700	16%	3225	0%
	Q25_s5_f0.5	3473	2700	22%	3275	5%
Input Data	none	3487	2540	27%	3138	10%

Note:

s: sliding window size.

Q: quality score average in the respecting window size.

(a) Quality Cutoff _90: the sequences after trimming by sliding window were checked again by Quality Cutoff performed by Galaxy (263). Two quality cutoff values were performed Q25_90 and Q20_90.

- Q25_90 means 90% of bases of a sequence have quality scores ≥ 25.
- Q20_90 means 90% of bases of a sequence have quality scores ≥ 20.

- Each Quality Cutoff contains 2 columns: numbers of sequences were retained and the percentage of bad sequences were removed after different quality score trimming parameters.

Condetri Q25_s5_f0.8 gave the lowest numbers of output sequences among parameters 3251/3487 input sequences. Other parameters yielded similar numbers of output sequences compared to input sequences (≥ 3473/ 3487).

Input data had the highest percentage of sequences removed with Quality Cutoff 90, 27% and 10% with Q25 and Q20, respectively. The results in Table 3.18 showed that quality trimming by Mothur with Q25_s5; Q35_s50 and Q25_s25 yeilded the better quality of sequences. After trimming, Mothur with Q25_s5; Q35_s50 and Q25_s25 gave higher quality of sequences with 2%, 4% and 10% of bad sequences removed, checking with Quality Cutoff 90%-Q25, respectively. Similarly, for checking with Quality Cutoff 90%-Q20, none of the putative bad sequences were removed with Mothur parameters Q25_s5 and Q35_s50, respectively; and just 2% of putative bad sequences removed with parameters and Q25_s25.

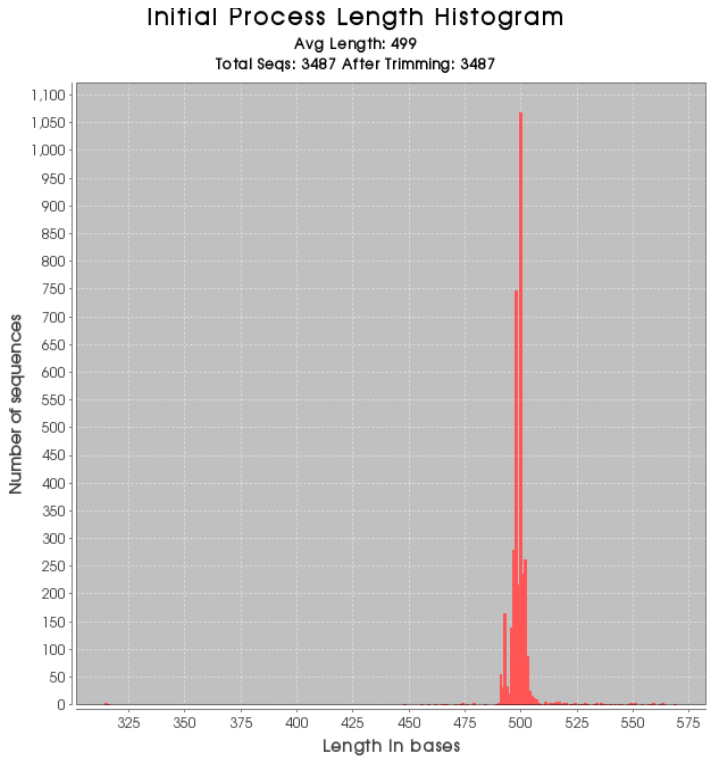
Length distribution and Average Quality Score distribution were surveyed for each parameter by RDP and Galaxy websites (**256**, **260**). Red peaks in Length distribution histogram generated by RDP website represent the number of sequences according to the y-axis and the length of sequences according to the x-axis (left graphs of **Fig. 3.9**). Yellow columns in the Average Quality Score distribution represent the average quality score for the whole sequences data according to y-axis and position of the bases according to the x-axis (right graphs of **Fig. 3.9**).

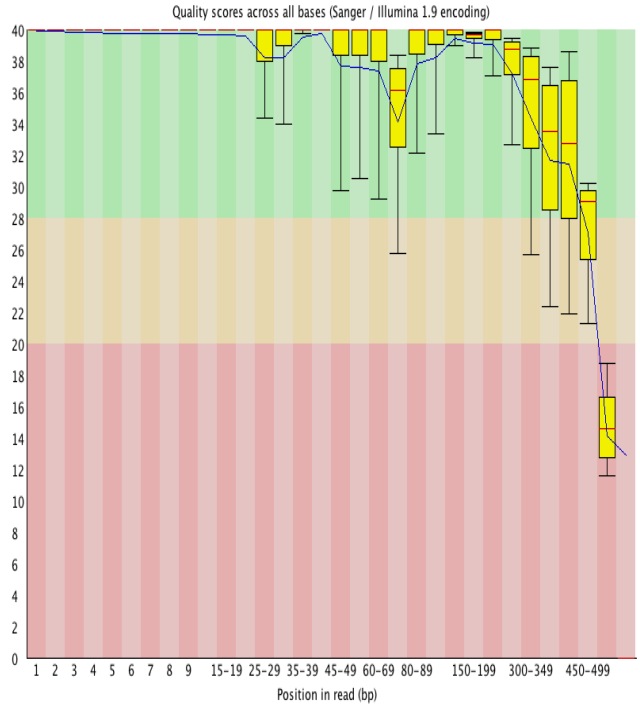
The results displayed in Figure 3.9 showed the differences of Length distribution and Average Quality Score of the sequences before and after going through different parameters of the trimming process (**Table 3.18**). Before quality score trimming (input sequence), the majority of the sequences had lengths of approximately 500 nt, with the average quality score of base positions 450-499 were generally < 20. Sequence trimming with Mothur Q25_s5, Q35_s50 had the best visualization with average quality score of bases position ≥ 28. However, the length of the sequences varied from 50 nt to 500 nt. It has been shown that classification accuracy is reduced with shorter amplicons (**335**, **336**). For this reason, although Mothur Q25_s5 and Q35_s50 gave the best quality score of the parameters, it could affect the downstream analyses.

Another parameter that gave better quality score sequences was Condetri Q25_s5_f0.8 with 16% putative bad sequences removed. Based on the results in Figure

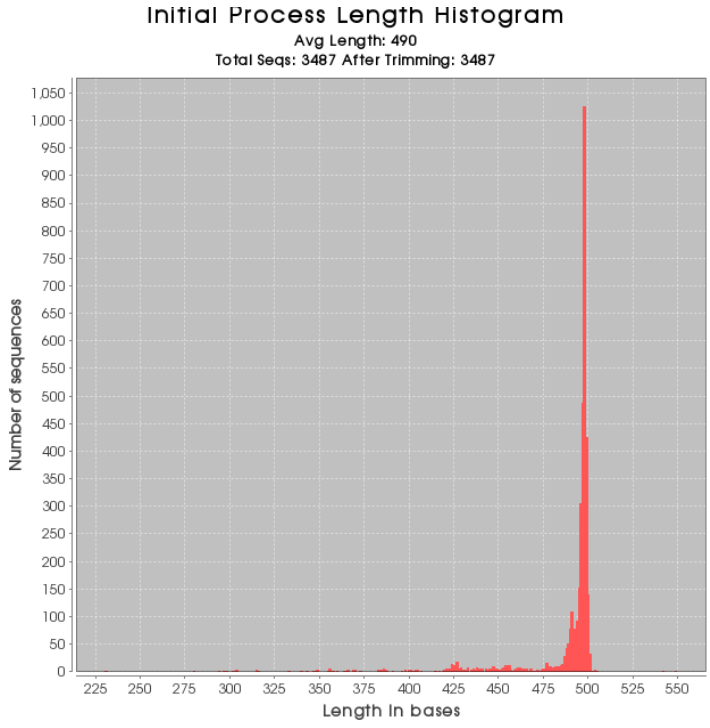
3.9, Length distribution and Average Quality Score of the sequences giving by the sequences trimmed by Condetri Q25_s5_f0.8 and Q25_s5_f0.5 were better than those of other parameters. The majority of sequence lengths were about 500 nt and the Average Quality Score of the sequences ≥ 20 for all the base positions with these parameters. Other parameters, including Galaxy Q25_s5, Q25_s10, Q25_s25, Mothur Q25_s25, Q25_s50 produced sequences with lower quality score (≤ 20) at positions 450-499 nt. Based on the data, Condetri Q25_s5_f0.8 and Q25_s5_f0.5 was chosen for quality score trimming of 454 sequencing data.

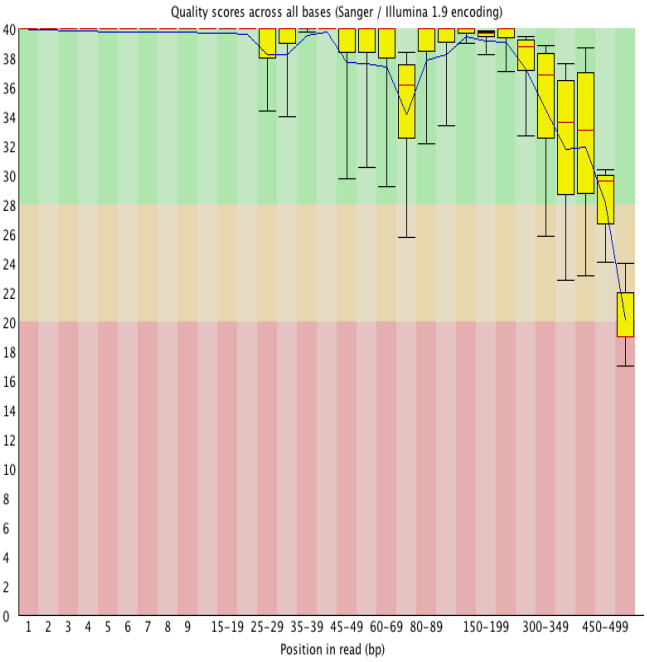
Input sequence



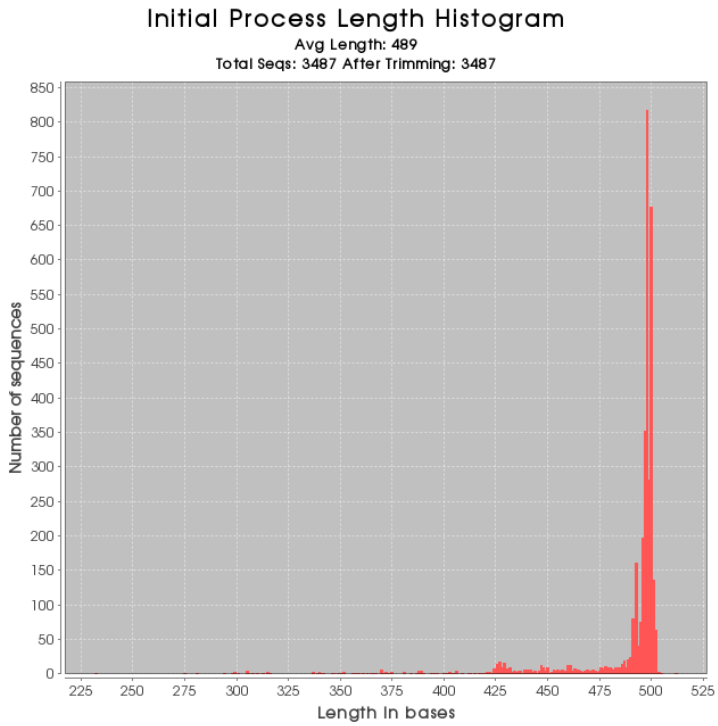


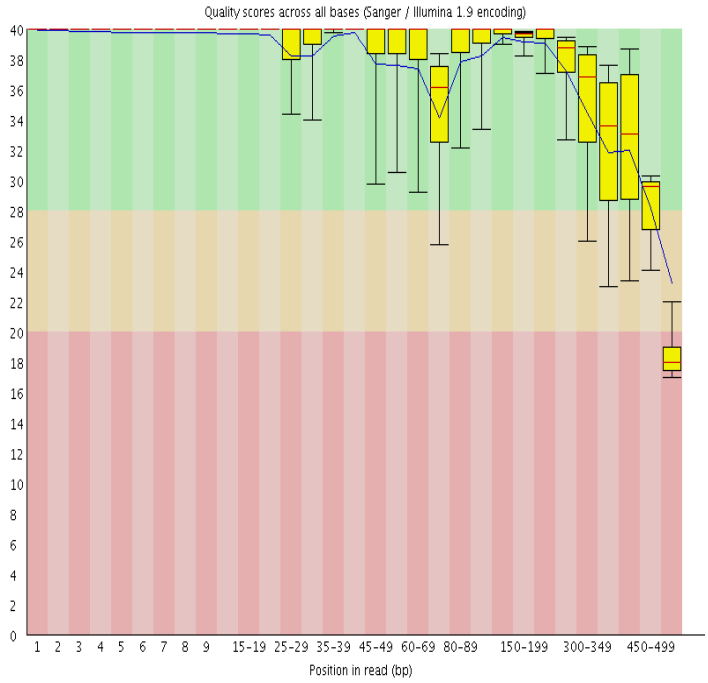
Galaxy Q25_s5



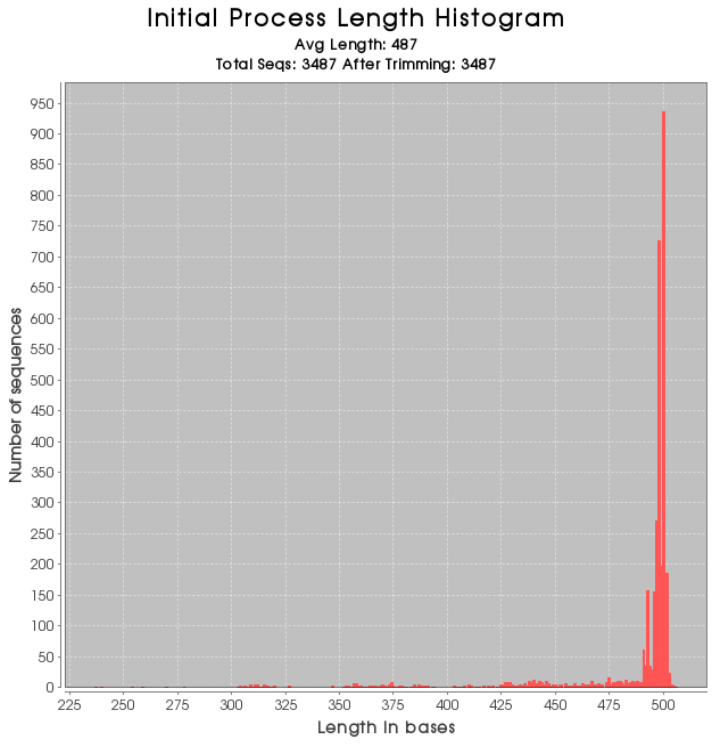


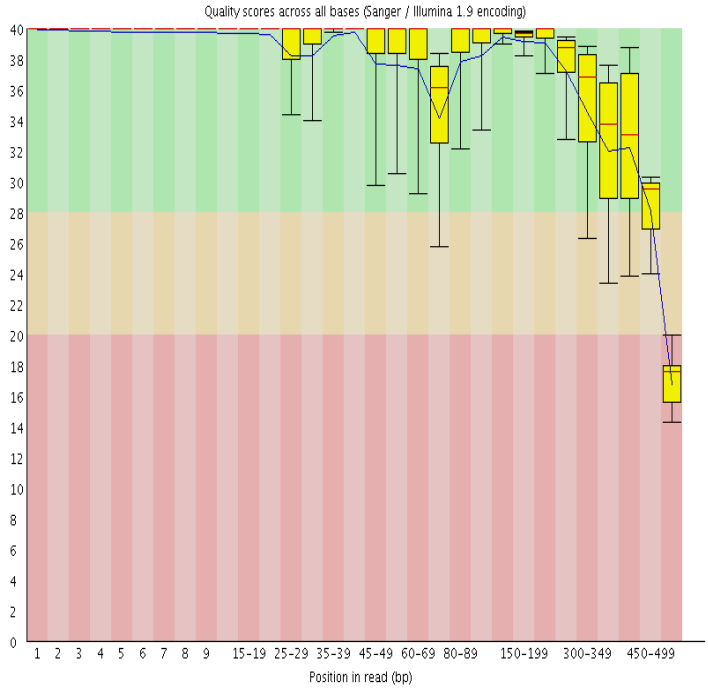
Galaxy Q25_s10



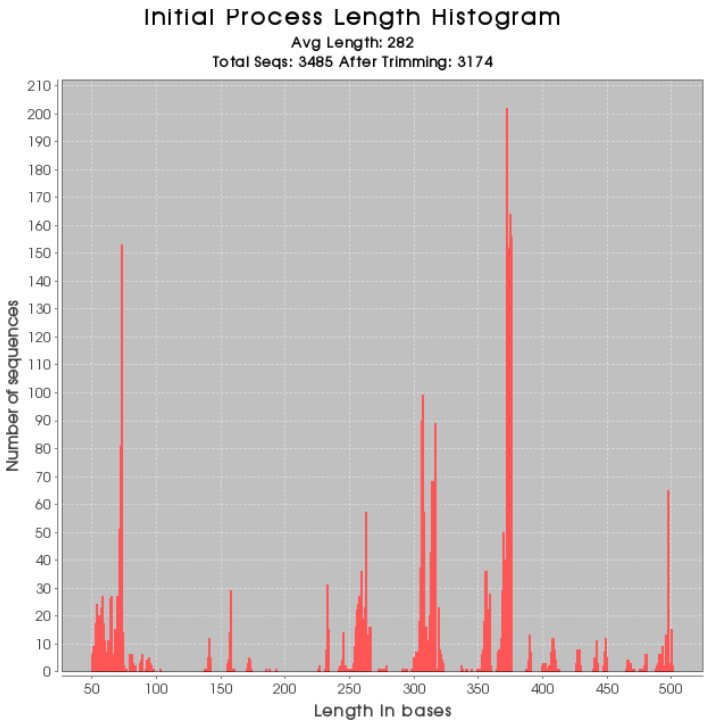


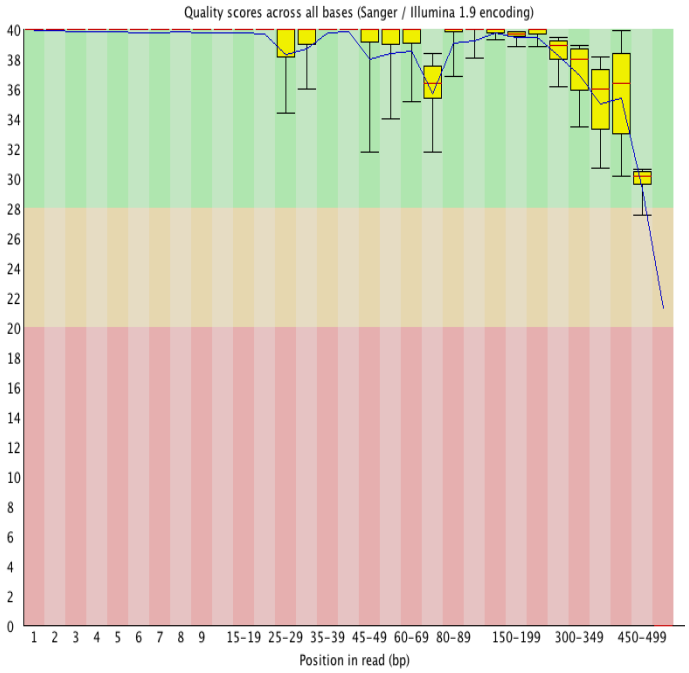
Galaxy Q25_s25



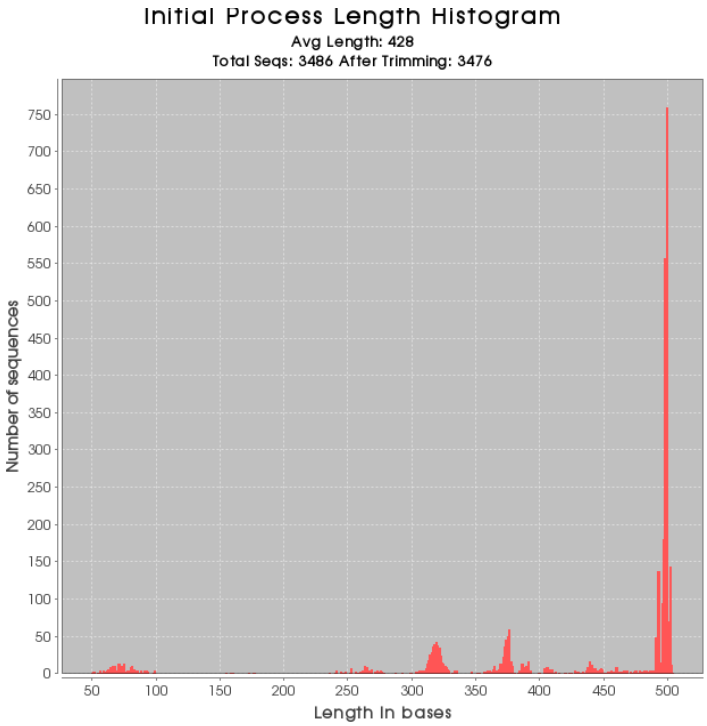


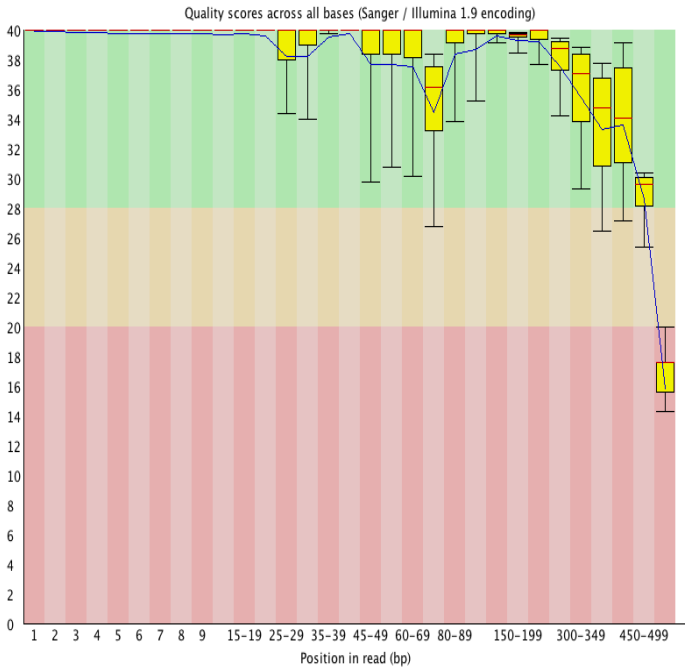
Mothur Q25_s5



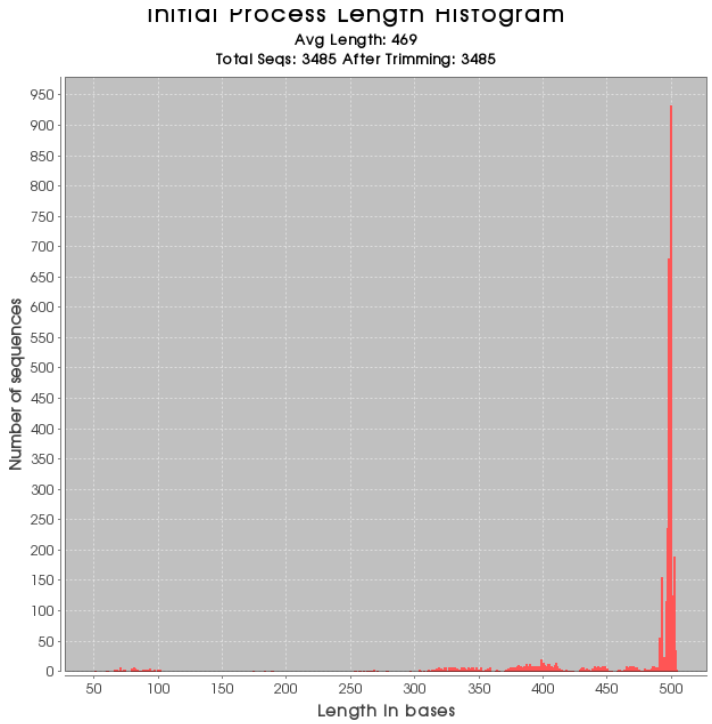


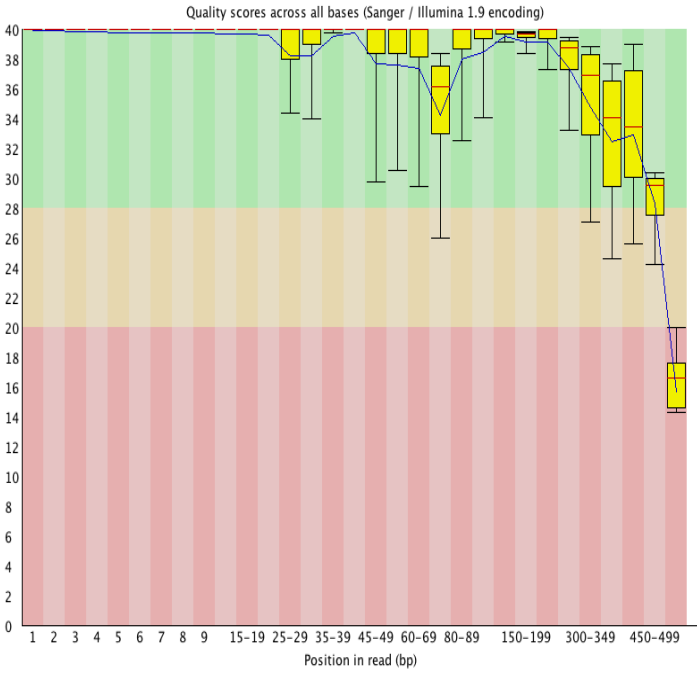
Mothur Q25_s25



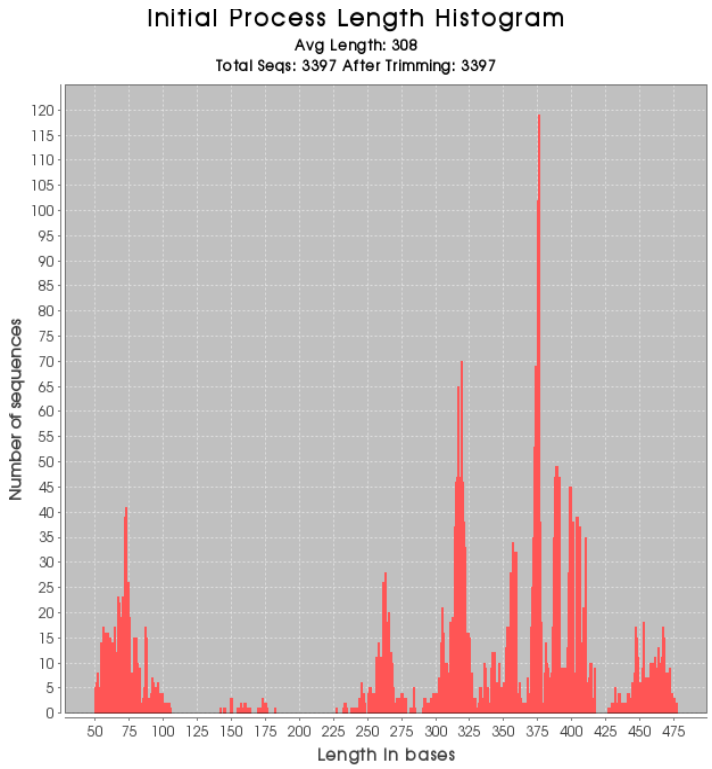


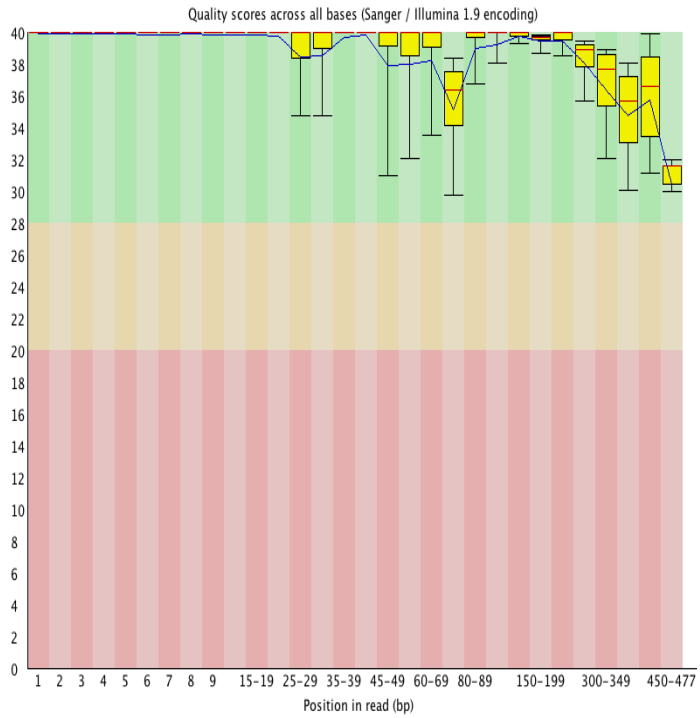
Mothur Q25_s50



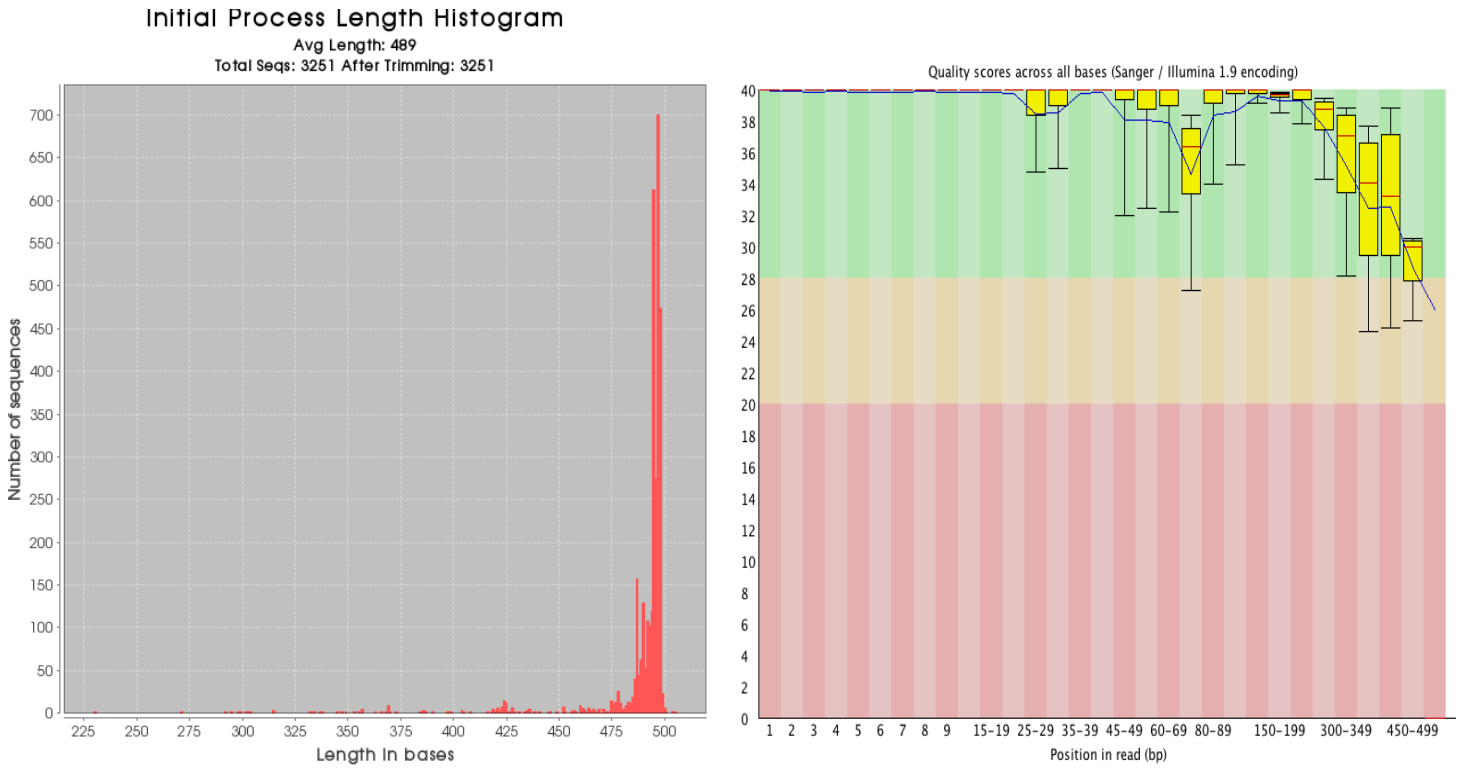


Mothur Q35_s50





Condetri Q25_f0.8



Condetri Q25/f0.5

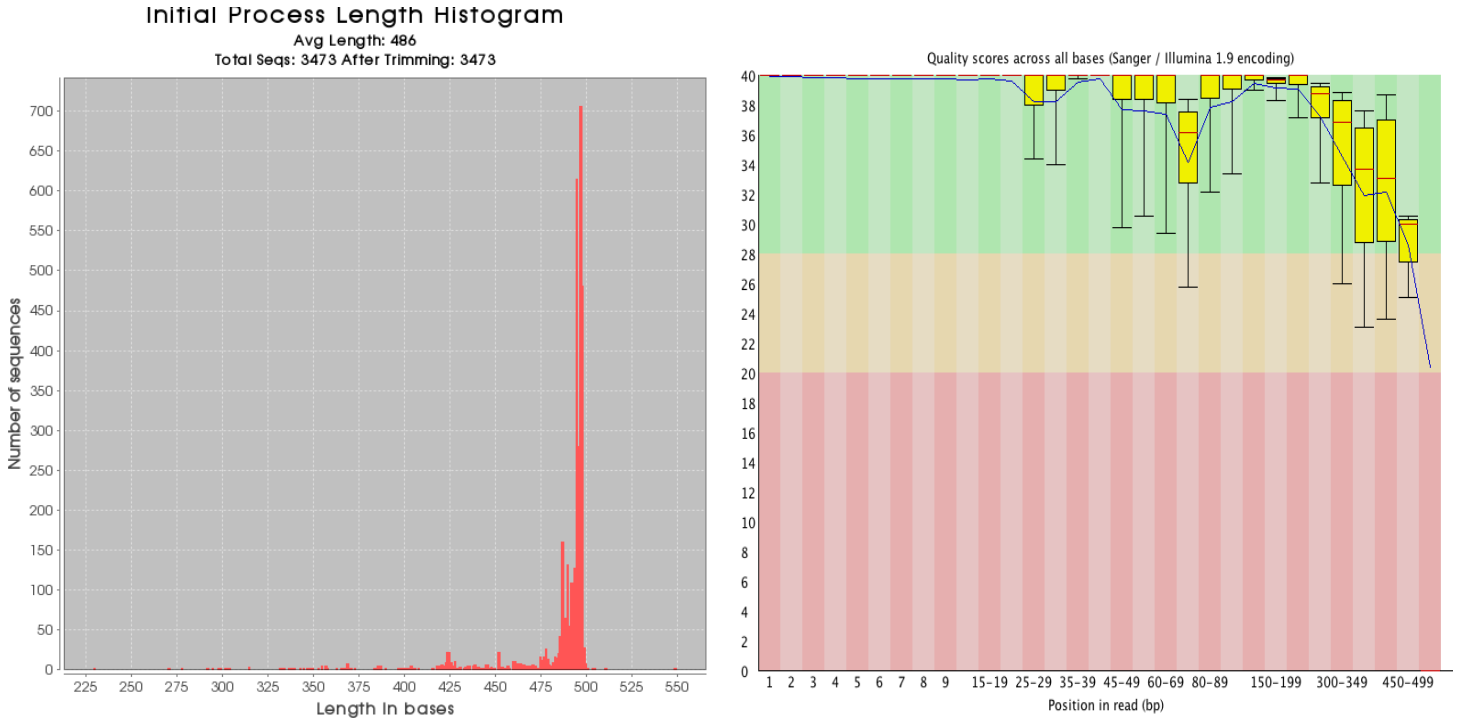


Figure 3.9. Length Distribution (left) analyzed by RDP program and Quality Score per base (right) analyzed by FastQC in Galaxy (**260**) for each parameters in Table 3.20 (**256**, **260**).

In order to optimize Condetri program, different parameters were tested using the same Input file described in Table 3.17.

3.1.3.3. Condetri trimming sequence algorithm:

The trimming of Condetri is performed in two steps:

(1) Trimming low quality bases from the 3'-end

(2) Overall quality check of read

Principle of Condetri program are shown in Figure 3.10, with the parameters (**243**):

-hq Hiqh quality threshold (Q_H)

-frac Fraction of read that must exceed (Q_H)

-minlen Min allowed read length

-mh When this number of consecutive Q_H bases (n_H is reached, the trimming stops)

-ml Max number of bases that have quality score < Q_H allowed after a stretch of hq bases from 3'-end.

-lq Low quality threshold.

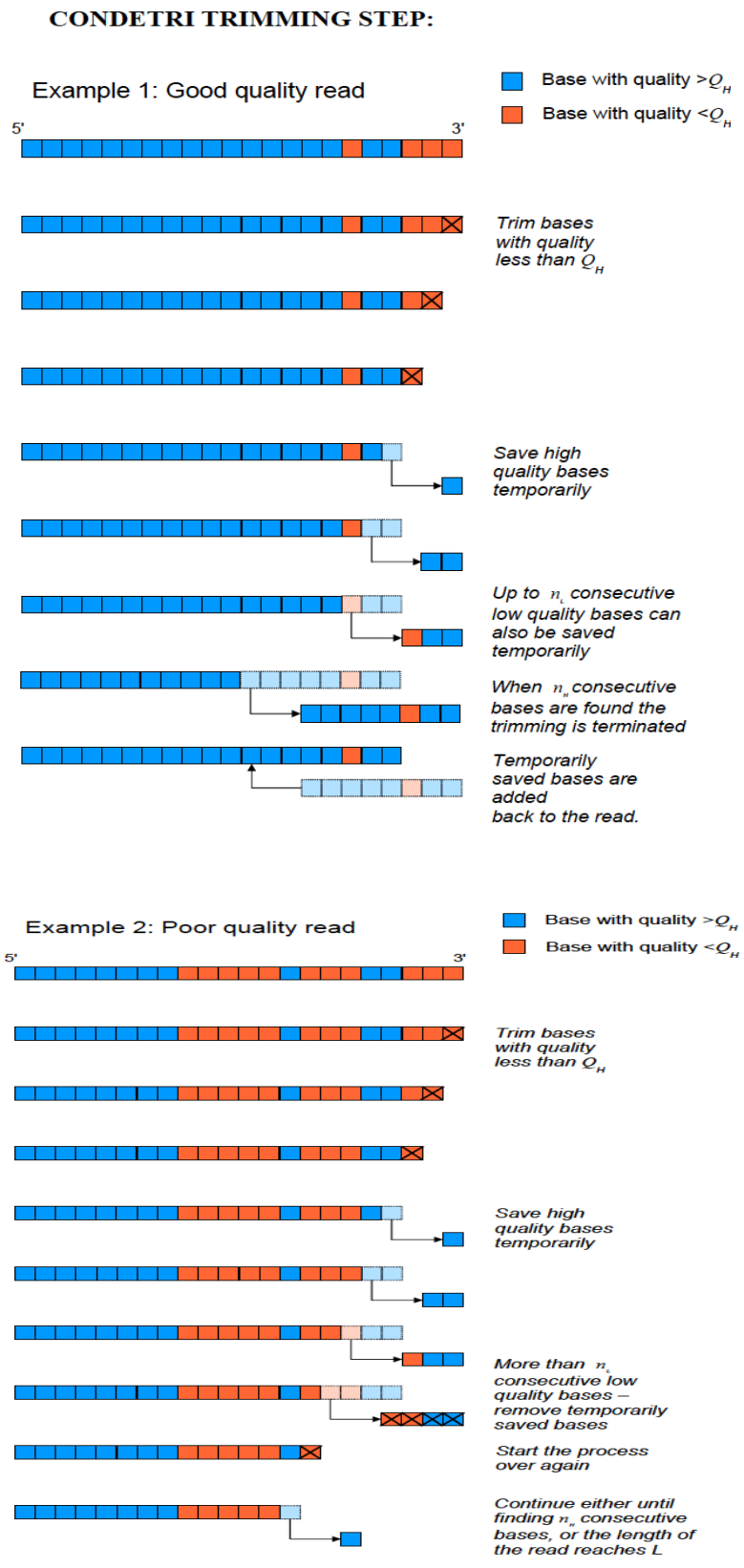


Figure 3.10. Trimming algorithm of Condetri software. Two examples of Good and Bad quality read trimmed by Condetri (**243**).

3.1.3.4. Optimizing Condetri parameters:

Purpose: to retain the highest numbers of sequences that have 90% Quality Score ≥ 25.

Table 3.19: Results of optimizing Condetri.

		-rmN	-hq	-lq	-frac	-lfrac	-minlen	-mh	-ml	Number of output reads	Number of output bases	Quality cutoff filtering_90%	
												Q25	Q20
Step 1	Default	V	25	10	0.8	0	50	5	1	3251 (93.23%)	1,590,487	2700 (16%)	3225 (0%)
		X	25	10	0.8	0	50	5	1	3251 (93.23%)	1,590,487		
Step 2	-mh-ml adjust	V	25	10	0.8	0	50	10	5	3324 (95.33%)	1,611,090		
		V	25	10	0.8	0	50	10	3	3326 (95.38%)	1,610,471		
		V	25	10	0.8	0	50	10	1	3328 (95.44%)	1,610,378		
Step 3	-frac adjust	V	25	10	0.9	0	50	10	5	2722 (78.06%)	1,331,706		
		V	25	10	0.9	0	50	10	3	2729 (78.26%)	1,334,134		
		V	25	10	0.9	0	50	10	1	2736 (78.46%)	1,334,940	2736 (0%)	2736 (0%)
Initial Input data	3487										1,740,625	2540 (27%)	3138 (10%)

Step 1: using default

Parameter checked: **-rmN**

Reason: -rmN is the removing of the non-ATCG characters from 5'-end before trimming, so that we should check with –rmN parameter and without parameter.

Result: same number of output reads and number of output bases.

Step 2: adjust the**–mh –ml** parameter

Parameter checked: -mh=10, -ml=5, 3 and 1 (respectively).

Reason:

-mh: When this number of consecutive Q_H bases (n_H) is reached, the trimming stops → -mh=10 –hq (Q_H)=25 means when 10 base that have the high quality score ≥ 25 → the trimming is stop (example 1 of Fig. 3.16).

-ml: Max number of bases that have quality score < Q_H allowed after a stretch of bases ≥ Q_H from 3'-end →–ml=3 –hq=25 means when 3 bases that have the quality score < 25 → the whole stretch of saved based were eliminate (example 2 of Fig. 3.16)

Result:

By changing the -mh=5 to the -mh=10, the numbers of output sequences and output bases increased from 93.23% to 95.33% and from 1,590,487 to 1,611,090, respectively. With -mh=10, number of output reads and number of output bases varied slightly among the –ml paramters (5, 3, 1), from 95.33% to 95.44% and from 1,611,090 to 1,610,378, respectively. However,

The -mh=10 gave more reads than –mh=5, but after checking with Quality filtering on Galaxy Q25p90 and Q20p90, small percent (4%-20%) low quality score bases left (data not show). For that reason, parameter –f of Condetri were adjusted from –f-0.8 to –f-0.9.

Step 3 : adjust **–frac0.9**

Parameter checked: -frac 0.9

Reason: Fraction of read that must exceed (Q_H) → -frac 0.9 hq=25 means read after trimming from 3’ end is checked again so that 90% of read have Q ≥ 25

Results: -mh=10, -ml=1 give the highest number of output reads and output bases.

The output fastq files were checked again with:

-Length Distribution and Average Quality Score per base. Results: OK (data not show).

-Quality cutoff Q25 and Q20 (Galaxy) have 0% of bad reads removed.

Summary, Condetri with parameters: -rmN Y, -hq 25, -lq 10, -frac 0.9, -lfrac 0, -minlen 50, mh=10 and –ml=1, were chosen for Quality Score trimming.

In Summary*:*

Based on the pipeline analyses, raw sequences with amplicon data type were trimmed by these criteria:

- Length (200-600 b)
- Homopolymer (8b)
- Split MID (0 error)
- Forward Primer Match (0 error)
- Quality trimming: Condetri
- Ambiguous base (N=0)

3.1.3.5. Discussion:

Mothur sliding window trimming has been used in many publications (**240**). However, there is no survey about Length Distribution of the data using this program. The variety of Length Distribution produced by Mothur should be assessed. Previous studies, such as Bowen et al. (2012), tested different quality score trimming parameters using the cutoff algorithm (**355**). The disadvantage of this algorithm is that there will be many sequences removed if they do not pass the cut-off value. Since the errors of sequences often occur at the end of the pyrosequencing process, removing bad quality bases at the 3' end of the sequences can help to save more sequences with quality as good as the cutoff algorithm.

In addition, we found that if sequences are removed due to ambiguous bases before quality trimming, about 20-40% of the sequences are lost. By Quality trimming using Condetri, the ambiguous nucleotides (assigned with Q=0) were removed at the 3' end of the sequences, helping to save more sequences in the trimming process.

3.1.4. Cut Adaptor Survey:

One of the factors that can affect the analyses of 16S rDNA sequences is the presence of sequences such as adaptors from the sequencing process (**356**)**.** The 454 pyrosequencing process uses two artificial sequences called Adaptor A and Adaptor B. Therefore, eliminating the Adaptors should be performed in the trimming process of raw data from 454 pyrosequencing. Raw sequences generated from 454 pyrosequencing contain only an Adaptor at the 3' end of the reads (**Fig. 3.12**). To remove these sequences, Cutadapt version 1.1 (http://galaxy.igmors.u-psud.fr/) was chosen.

3.1.4.1. Principle of the CutAdaptor version 1.1 (information from the website):

Cutadapt correctly deals with partial adapter matches. Based on its principle, if we increase the error rates of the Cut Adaptor, we increase the chance to detect the Adapter from the 3' end of the sequences. For 454 sequencing, the Quality of the Base read decreases at the end of the sequencing process. This means that at the end of the read, we fewer true sequences present in the sample. In addition, 454 sequencing is based on light signal reading, and sometimes a base is overcalled (insertion) or undercalled (deletion). The Adaptor A and the Primer 27F present at the 3' end of the read are in positions that sequencing errors occur frequently, so that we do not expect to find the perfect match of Adaptor A and Primer 27F. For this reason, in order to improve the Cut Adaptor efficiency, we should allow higher error rates of the program.

3.1.4.2. Optimizing CutAdaptor version 1.1:

In order to find which is the best parameters to trim the Adaptor A and Primer 27F at the 3' end of the sequences, a data of 149 sequences that have a minimum length of 526 nucleotides were input into the program. The 27F primer (CTGAGCCAKGATCAAA-CTC, length19) was used 15%, 20%, 25%, 30% error rates of the CutAdaptor version 1.1. The results are in Table 3.22.

Table 3.20: Numbers of trimmed sequences and trimmed bases through different error rates.

Error rates	15%	20%	25%	30%
N^0 of trimmed sequences	24	38	58	83
N^0 of trimmed bases	0.52%	0.86%	1.35%	2.65%

The results in Table 3.20 showed that when the error rates were increased, the number of trimmed sequences and trimmed bases also increased. Notice that, just the

adaptors (or the primers) at the 3’ end of the read is trimmed but not the entire read. That means, if the input is 149 sequences, the output is still 149 sequences with the adaptors or primers are removed. Some of the sequences from the 15%, 20%, 25% and 30% error rate trimmings were extracted, using the alignment by both manual (by eyes) and the website Multalin version 5.4.1 (http://multalin.toulouse.inra.fr/multalin/cgi-bin/multalin.pl) in order to check the differences among these error rates. The results showed that the CutAdaptor version 1.1 performed a best a the 25% error rate and thus with a higher chance to trim the 3’ primers than the error rates of 15% and 20% due to its ability to allow 3 errors in the alignment (1 deletion, 1 insertion and 1 mismatch). An error rate 30% is too strong for the trimming since it removed too much non-specific sequences at the 3’end of the read (data not shown).

3.1.4.2. Optimizing CutAdaptor version 1.1:

The sequences from a sediment sample with 12,351 sequences, were random picked in order to test the Cutadapt version 1.1 with 4 parameters. The sequences file was put into the RDP program to obtain sequence length distribution (**Fig. 3.11**).

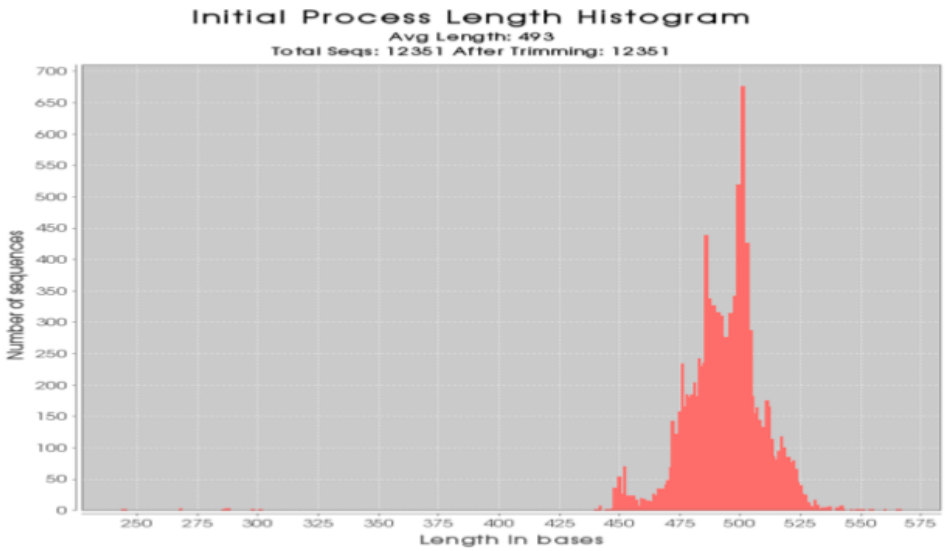


Figure 3.11. Length distribution of tested sequence plotted by RDP program**.**

Raw sequences schematic obtained from 454 pyrosequencing are presented in Figure 3.12. Sequences that were obtained from 454 pyrosequencing contained:

(i) 10 nt of MID (or Barcode) and primer 518R (17 nt) at the 5' end of the sequences.

(ii) The 16S rDNA region spanning from primer 518R, including V3, V2 and V1.

(iii) Complement sequences of primer 27F and adaptor A.

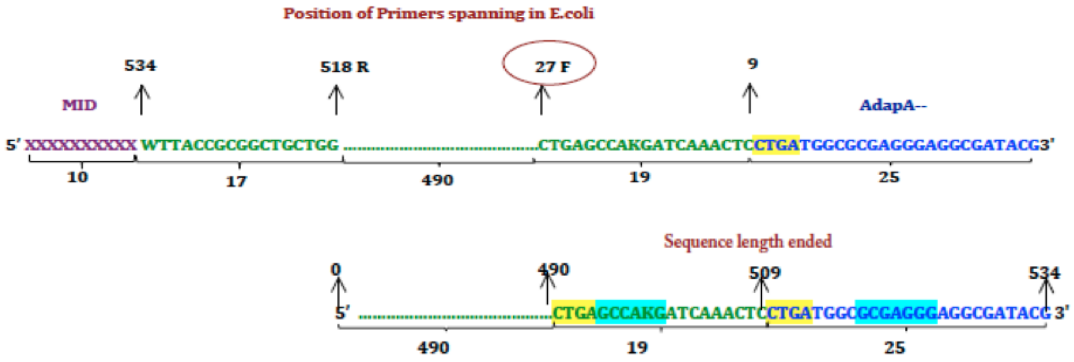


Figure 3.12. Raw sequence schematic with the position of adaptor A and primer 27F (based on the location of *E. coli* 16S rDNA sequence).

Note:

Upper figure: sequences before splitting MID and primer 518R, with the positions of the primers and their length. Sequences have 10 nt of MID and 17 nt of 518R primer at the 5'end. The length of 16S rDNA region that located between the primers should be in 490 nt length.

Lower figure: Sequences after splitting Mid and primer 518R, with the positions of the 16S rDNA region , complement sequences of 27F and adaptor A according to the length of the reads.

According to Figure 3.12, sequences that have length:

➢➢556---535: should contain the adaptor A, and primer 27F or something else at the 3' end.

➢➢534---510: should contain the adaptor A, and primer 27F at the 3' end.

➢➢509---491: should not contain the adaptor A, but contain the primer 27F at the 3' end.

➢➢490--- 244: should not contain the adaptor A neither the primer 27F at the 3' end.

The input file contains the sequences with a minimum length of 244 nt and a maximum length of 556 nt. The input files were split (using Mothur) into 4 length-range data sets according to the position of Adaptor A and Primer 27F in Figure 3.12 **(Table 3.21).**

Table 3.21: Four data sets according to their length range.

Length range	556-535	534-510	509-491	490-244
N^0 of sequence	41	1649	5751	4910

Four split files were then input in http://galaxy.igmors.u-psud.fr/ using Cut Adaptor version 1.1 with parameter maximum error rate 25.00%. For each range of the sequences, two Cut Adaptor tests were executed. One was for cutting adaptor A and another one was for cutting primer 27F. These two tests used the same input file and were performed separately. The purpose of this performance is to find out where is the Adaptor A and where is the Primer 27F in the sequences.

3.1.4.3. Results:

According to **Figure 3.12**, sequences that have length from:

➢➢ Dataset with 556---535 nt length should contain both adaptor A (full length of 25nt) and the primer 27F (full length of 19 nt).

The result showed that 7.3% of the sequences contained adaptor A and 18.5% of sequence contained primer 27F (**Table 3.22**).

Manual checking with:

a) Adaptor cutting test:

Sequences that were cleaved 4 nucleotides at the 3'end did not represent the adaptor A and primer 27F. One sequence that was cleaved 7 nucleotides at the 3' end did not represent the adaptor A.

b) Primer cutting test:

Sequences that were cleaved at 15, 17 and 19 nucleotides at the 3'end did represent primer 27F. With the 2 sequences were cleaved >=35 nucleotides at 3'end, one with 58 nucleotides removed was adaptor A and primer 27F and another one with 35 nucleotides was primer 27F and a stretch of oligonucleotides.

Dataset with 556--- 535 nt with 41 seqs should contain both adaptor A and primer 27F. However, manual checking with the sequences that were cut from the program, just 1 sequences that contained both adaptor A and primer 27F were found.

➢➢ Dataset with 534--- 510 length should contain both adaptor A (length from 0 to 25 nt) and the primer 27F (full length of 19 nt).

The showed 15.9% of the sequences contained adaptor A and 84.3% of sequences contained primer 27F (**Table 3.22**).

Manual checking with:

a) Adaptor cutting test:

In the Adaptor cutting test, sequences that were removed from nucleotides f3 to 7 at the 3'end did not really represent the adaptor A. Sequences that were removed from nucleotides 8 to 25 did represent the adaptor A (manual checking) (**Table 3.22**). The number of sequences in the 534--- 510 length dataset is 1649, with just 16 sequences (~1%) containing the adaptor A.

Table 3.22: Length cutting for four lengths ranges in adaptor A & primer 27F trimmings:

Length cutting	Adaptor				Primer			
	556-535	534-510	509-491	490-244	556-535	534-510	509-491	490-244
3	0	9	31	20	0	6	21	11
4	2	60	122	21	2	13	44	15
5	0	15	54	11	0	2	9	2
6	0	28	128	72	0	31	145	74
7	1	134	440	115	0	145	451	119
8	0	7	3	1	0	35	50	10
9	0	1	0	0	0	5	15	3
10	0	1	0	0	0	22	60	6
11	0	0	0	0	0	10	18	2
12	0	4	2	0	0	44	128	6
13	0	0	0	0	0	9	19	1
14	0	1	0	0	0	7	10	1
15	0	0	0	0	1	84	265	22
16	0	0	0	0	0	52	95	33
17	0	1	0	0	2	528	2533	3211
18	0	0	0	0	0	175	295	363
19	0	0	0	0	1	108	195	152
20	0	0	0	0	0	56	70	53
21	0	0	0	0	0	30	25	10
22	0	0	0	0	0	2	7	0
23	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0
25	0	1	0	1	0	0	0	0
>=23	0	0	0	0	0	26	19	0
>=24	0	0	0	0	0	0	0	3
>=35	0	0	0	0	2	0	0	0
>=206	0	0	1	0	0	0	0	0
Total N°trimmed sequences	3	262	781	241	8	1390	4474	4097
% of trimmed sequences	7.3%	15.9%	13.6%	4.9%	19.5%	84.3%	77.8%	83.4%
% of trimmed bases	0.07%	0.19%	0.17%	0.06%	0.76%	2.54%	2.36%	2.89%

Note:

- Length cutting 3 means that 3 bases are trimmed from the 3' end of the read.
- Length cutting 20 means that 20 bases are trimmed from the 3'end of the read.
- In the 556-535 length range, length cutting 4 and number of the sequences according of this

length is 2 mean there are two sequences that have four bases trimmed at the 3'end.

b) Primer cutting test:

In the Primer cutting test, sequences that were cleaved from nucleotides 3 to 6 at the 3'end did not represent the primer 27F. Sequences that were cleaved from nucleotides 7 to 22 did represent the primer 27F and consist of ~ 80% of total number of sequences (1649 seqs).

Although the length of primer 27F is 19 nt, the 27 F sequence at the 3' end of the reads are not exactly the same due to the errors (indels and mismatches) generated at the end of pyrosequencing process. Hence, the length of the removed sequences can vary from 16 nt to 22nt. The sequences that have length ≥23 contained Primer 27F and Adaptor A, or the error sequence behind.

➢➢ Dataset with 509--491 nt length should contain no adaptor A, but contain primer 27F (length from 0 nt-19 nt).

The results showed that 13.6% of the sequences contained adaptor A, and 77.8% of the sequences contained primer 27F (**Table 3.22**).

Manual checking with:

a) Adaptor cutting test:

Sequences that were cleaved from nucleotides 3 to 8 at the 3'end did not represent the adaptor A, just 1 sequence (length cutting 12) contained the adaptor A and 1 sequence with length cutting >=206) (**Table 3.22**) (checking by eyes).

b) Primer cutting test:

Sequences that were cleaved from nucleotides 3 to 6 at the 3'end did not represent the primer 27F. Sequences that were cleaved from nucleotides 7 to 22 did represent the primer 27F and consist of ~ 74% of total number of sequences (5751 seqs) (**Table 3.22**).

➢➢ Dataset with 490--244 nt length should contain no adaptor A and no primer 27F.

The results showed that 4.9% of the sequence contained the adaptor A and 83.4% of the sequence contained the primer 27F.

Manual checking with:

a) Adaptor cutting test:

Sequences that were cleaved from nucleotides 3 to 8 at the 3'end did not represent the adaptor A, just 1 sequence with length cutting 25 nt contained the adaptor A.

b) Primer cutting test:

Sequences that were cleaved from nucleotides 3 to 6 did not represent the primer 27F. Sequences that were cleaved from nucleotides 7 to 22 did contain the primer 27F and

consist of ~ 81.3% of the total number of sequences (4910 seqs) (**Table 3.22**).

However, there should be no primers in this length range according to **Figure 3.12**. Possibly, the V3-V1 amplicon of these sequences were shorter than that of *E. coli*. These sequences that had length cutting >=24, 1 contained the adaptor A and primer 27F, 2 other sequences contained adaptor A and primer 27F and some nucleotides after that.

3.1.4.4. Discussion:

Four datasets with different length range were optimized the determination of the CutAdapt version 1.1 in order to see positions of adaptor A and primer 27F compared with the standard *E. coli* position (**Fig. 3.13**). The results showed that the V3-V1 amplicon in the environment can be different compared to the standard positions of *E. coli*. Through the manual checking procedure of the CutAdapt process, pyrosequencing errors at the 3' end of the sequence were indentified.

3.2. Chemical analyses of the SG-DN river system:

3.2.1. Data obtained in February 2012:

3.2.1.1. Total Organic Carbon (TOC):

Total organic carbon (TOC) is the amount of carbon found in organic compounds and is often used as a non-specific indicator of water quality (**264**). The TOC, together with total nitrogen (TN) contents in soils and sediments, are the important parameters in the environmental status estimation. They are mainly derived from decomposition of the plants and animals or plankton or from anthropogenic sources such as chemical contaminants, fertilizers or organic rich waste. The sediments and soil concentrations of the organic carbon are well correlated with organic contaminants, and for this reason they can be used as a tool for the estimation of the level of contamination toxicity.

TOC levels of the sediment samples from 8 locations are from 3.1 to 5.1 % (**Table 3.23**). Among all the samples, SG1 and SG3 have the highest TOC level (5.1 %), while sample from SG2 had the lowest TOC level (3.1%). The average of TOC levels in the SaiGon river is 4.5%. In the DongNai river, the TOC level of DN2 is highest (4.6%) and that of DN1 is lowest (3.9%). The average TOC level of all the samples is 4.36 %. The TOC level of the sample from the intersection (SG5) is similar to that of the average (4.4%/4.36 %).

3.2.1.2. Heavy Metals:

Table 3.23 discribes the chemical analyses of 8 sediment samples from 8 locations taken on February 2012, the average and the TEL & PEL value for each chemical. They are Total Organic Carbon (TOC), seven heavy metals (Pb, Cu, Ni, Cr, Zn, Hg, Cd), Total PCBs and Total PAHs. Among the heavy metals, Hg and Cd are not detected with the threshold < 1 $\mu g.kg^{-1}$ dry weight in all sediment samples.

Pb concentrations in the sediment samples from 8 locations are from 17.2-43.4 $mg.kg^{-1}$, below the PEL (Table 3.2). The concentration of Pb are highest in SG3 (43.4 $mg.kg^{-1}$), lies between TEL and PEL, and lowest in SG2 (17.2 $mg.kg^{-1}$), below TEL. In the DongNai river, Pb concentration of DN1 are highest (30.3 $mg.kg^{-1}$), above TEL, and lowest in DN2 (26.1 $mg.kg^{-1}$), below TEL. The average Pb concentration of all the samples is 29.3 $mg.kg^{-1}$ and do not exceed TEL. The Pb concentration of the intersection is higher than the TEL value (32.3 $mg.kg^{-1}$).

The concentrations of Cu in the sediment samples from 8 locations are from 13.9 to 57.9 $mg.kg^{-1}$, below the PEL. The concentrations of Cu in SG3 are highest (57.9 $mg.kg^{-1}$

^{1}), lies between TEL and PEL, and lowest in SG2 (13.9 mg.kg^{-1}), below TEL. In the DongNai river, Cu concentrations are quite similar among the samples from 24.4-27.6 mg.kg^{-1}. The average Cu concentration of all the samples is 32.2 mg.kg^{-1}, above the TEL value. The Cu concentration of the intersection is 45.3 mg.kg^{-1}, above the TEL value.

The Ni concentrations of the sediment samples from 8 locations are from 22.4 to 83.2 mg.kg^{-1}. The Ni concentrations in SG5 are highest (83.2 mg.kg^{-1}) and lowest in SG2 (22.4 mg.kg^{-1}). In the DongNai river, Ni concentrations are quite similar among the samples from 49.8 to 59.3 mg.kg^{-1}. The average Ni concentration of all the samples is 52.8 mg.kg^{-1}. There is no TEL and PEL for Ni. However, according to the Ontario (Canada) Ministry of Environment Screening Level Guidelines, if the Ni concentrations is over 75 mg.kg^{-1} dry weight, it could be a problem (**265**). The Ni concentration in SG5 sample is above this value.

The Cr concentrations of the sediment samples from 8 locations are from 21.1 to 54.5 mg.kg^{-1}, below the PEL level. The Cr concentrations in SG5 are highest (54.5 mg.kg^{-1}), lying between TEL and PEL, and lowest in SG2 (21.1 mg.kg^{-1}), below the TEL level. In the DongNai river, the Cr concentrations are quite similar among the samples from 41.0 to 49.2 mg.kg^{-1} and all below the TEL. The average Cr concentration of all the samples is 42.9 mg.kg^{-1}.

The Zn concentrations in the sediment samples from 8 locations are from 65.0 to 168.0 mg.kg^{-1}, all below the PEL. Samples from locations SG3, SG6, SG5, SG4 and DN2 are above the TEL level. The Zn concentrations in SG5 are highest (168.0 mg.kg^{-1}) and lowest in SG2 (65.0 mg.kg^{-1}) and below the TEL. In the DongNai river, the Zn concentrations are quite similar among the samples from 112.0-134.0 mg.kg^{-1}. The average Zn concentration of all the samples is 124.4 mg.kg^{-1}, above the TEL value.

Table 3.23: Chemical analysis of the SG-DN river system in February 2012.

River	Sample	TOC (% dry wt)	Pb	Cu	Ni	Cr	Zn	Hg	Cd	Total PCBs ($ng.g^{-1}$dry wt)
			$mg.kg^{-1}$							
SaiGon	SG1	5.1	28.8	22.2	42.0	35.6	73.4	ND	ND	0.44
	SG2	3.1	17.2	13.9	22.4	21.1	65.0	ND	ND	1.5
	SG3	5.1	43.4	57.9	51.6	48.2	163.0	ND	ND	1.7
	SG6	4.7	25.4	38.8	58.8	45.4	158.0	ND	ND	0.98
DongNai	DN1	3.9	32.0	24.4	55.0	41.0	112.0	ND	ND	1.8
	DN2	4.6	26.1	27.5	59.3	48.4	134.0	ND	ND	0.73
	SG4	4.0	29.0	27.6	49.8	49.2	122.0	ND	ND	1.04
Junction	SG5	4.4	32.3	45.3	83.2	54.5	168.0	ND	ND	0.48
Mean		4.4	29.3	32.2	52.8	42.9	124.4	0.13	0.7	1.08
TEL		N	30.2	18.7	N	52.3	124.0	0.13	0.7	21.5 [a]
PEL		N	112.0	108.0	N	160.0	271.0	0.7	4.2	189.0 [a]

Note:

- There is no criteria for heavy metals concentrations in sediments in Vietnam, so we will base on the ISQG (Interim Sediment Quality Guidelines) of Canadian Sediment Quality Guidelines, which has been used in previous studies of the SG-DN river, for the Protection of Aquatic Life – Update 2002 (**266**).

- ND = non detected with the LOD (Limit of Detection) is < 0.001 $mg.kg^{-1}$ dry weight.

- N : there is no ISQG value for this category, a : $mg.kg^{-1}$

- Based on ISQG, there are two Effects Level :

i/ TEL (Threshold Effects Level) : represents the concentration below which adverse effects are expected to occur only rarely.

ii/ PEL (Probable Effects Level) : represents the concentration above which adverse effects are frequently expected.

3.2.1.3. PAHs:

Table 3.24 describes the 13 PAH compounds analyses of the 8 sediment samples from 8 locations. Among these 13 PAHs compounds, Acenaphthylene, Acenaphthene and Phenanthrene were not detected with the threshold < 1 $ng.g^{-1}$ dry wt in all the sediment samples.

Naphthalene:

The naphthalene concentrations of the sediment samples from 8 locations are from 25.0 to 133.0 $ng.g^{-1}$ below the PEL level. The concentrations of naphthalene in SG1 are highest (133.0 $ng.g^{-1}$), lying between TEL and PEL, and that of DN2 is lowest (25.0

ng.g^{-1}), and below TEL. In the DongNai river, the naphthalene concentrations are quite similar among the samples from 25.0 to 33.0 ng.g^{-1} and all below TEL. The average naphthalene concentration of all the samples is 34.6 ng.g^{-1}. The naphthalene concentration of the intersection is 36.0 ng.g^{-1}, above the TEL value.

Anthracene :

The anthracene concentrations of the sediment samples from 8 locations are from 36.0 to 65.0 ng.g^{-1}, below the PEL level. The concentration of anthracene in SG2 is highest (65.0 ng.g^{-1}), lying between TEL and PEL, and that of SG5 is lowest (36.0 ng.g^{-1}), and below TEL. In the DongNai river, the anthracene concentrations are highest in DN2 (52.0 ng.g^{-1}), above TEL, and lowest in DN1 (37.0 ng.g^{-1}), below TEL. The average anthracene concentration of all the samples is 46.8 ng.g^{-1}.

Fluoranthene :

The fluoranthene concentrations of the sediment samples from 8 locations are from 28.0 to 54.0 ng.g^{-1} and all below the TEL level. The concentration of fluoranthene are highest in SG4 (54.0 ng.g^{-1}) and lowest in DN1 & DN2 (28.0 ng.g^{-1}). In the SaiGon river, the fluoranthene concentrations in SG3 are highest (52.0 ng.g^{-1}) and lowest in SG1 (38.0 ng.g^{-1}). The average fluoranthene concentration of all the samples is 39.0 ng.g^{-1}. The fluoranthene concentration of the intersection SG5 is 30.0 ng.g^{-1}, similar to that of average.

Pyrene :

The pyrene concentrations of the sediment samples from 8 locations are from 28.0 to 69.0 ng.g^{-1}, and all below the TEL level. The concentrations of pyrene are highest in SG3 (69.0 ng.g^{-1}) and lowest in SG1 (28.0 ng.g^{-1}). In the DongNai river, the pyrene concentrations are highest in SG4 (67.0 ng.g^{-1}) and lowest in DN2 (29.0 ng.g^{-1}). The average pyrene concentration of all the samples is 42.3 ng.g^{-1}. The concentration of the pyrene of the intersection SG5 is 40.0 ng.g^{-1}, similar to that of average.

Benzo[a]pyrene :

The benzo[a]pyrene concentrations of the sediment samples from 8 locations are from 20.0 to 43.0 ng.g^{-1} and all below the TEL level. The concentrations of benzo[a]pyrene are highest in SG4 (43.0 ng.g^{-1}) and lowest in DN2 (20.0 ng.g^{-1}). In the SaiGon river, the benzo[a]pyrene concentrations are quite similar among the samples SG1, SG2 and SG6 (22.0-23.0 ng.g^{-1}), and highest in SG3 (33.0 ng.g^{-1}). The average benzo[a]pyrene concentrations of all the samples is 26.1 ng.g^{-1}. The concentration benzo[a]pyrene of the intersection SG5 is 23.0 ng.g^{-1}, similar to that of average.

Dibenz(a,h)anthracene :

The dibenz(a,h)anthracene concentrations of the sediment samples from 8 locations are from 17.0 to 35.0 ng.g^{-1} and all above the TEL level. The concentrations of dibenz(a,h)anthracene are highest in SG1 (35.0 ng.g^{-1}) and lowest in DN1 (17.0 ng.g^{-1}). In the SaiGon river, dibenz(a,h)anthracene concentration is highest in SG1 (35.0 ng.g^{-1}) and lowest in SG6 (19.0 ng.g^{-1}). In the DongNai river, the dibenz(a,h)anthracene concentration are quite similar among the samples from 17.0 to 22.0 ng.g^{-1}. The average dibenz(a,h)anthracene concentration of all the samples is 22.4 ng.g^{-1}. The dibenz(a,h)anthracene concentration of the intersection is 18.0 ng.g^{-1}, similar level to samples SG6, DN1 and DN2. The average dibenz(a,h)anthracene concentrations of all the samples is 22.4 ng.g^{-1}.

Benzo[g,h,i]perylene :

The benzo[g,h,i]perylene concentrations of the sediment samples from 8 locations are from 19.0 to 65.0 ng.g^{-1}. The concentration of benzo[g,h,i]perylene are highest in SG4 (65.0 ng.g^{-1}) and lowest in DN2 (19.0 ng.g^{-1}). In the SaiGon river, benzo[g,h,i]perylene are highest in SG3 (59.0 ng.g^{-1}) and lowest in SG6(22.0 ng.g^{-1}). Other samples such as SG1 and SG2 have higher level of benzo[g,h,i]perylene than SG6 (31.0 & 29.0, respectively). In the DongNai river, the benzo[g,h,i]perylene concentrations are highest in SG4 (65.0 ng.g^{-1}) and lowest in DN2 (19.0 ng.g^{-1}). The benzo[g,h,i]perylene concentration in DN1 is slightly higher than that in DN2 (24.0 ng.g^{-1}). The average benzo[g,h,i]perylene concentration of all the samples is 34.5 ng.g^{-1}. The benzo[g,h,i]perylene concentration of the intersection is 27.0 ng.g^{-1}, lower than average concentration.

Indeno[1,2,3-cd]pyrene :

The indeno[1,2,3-cd]pyrene concentrations of the sediment samples from 8 locations are from 21.0 to 57.0 ng.g^{-1}. The concentrations of indeno[1,2,3-cd]pyrene are highest in SG4 (57.0 ng.g^{-1}) and lowest in DN2 all the samples (21.0 ng.g^{-1}). In the SaiGon river, the indeno[1,2,3-cd]pyrene concentrations are highest in SG3 (49.0 ng.g^{-1}) and is lowest in SG6 (23.0 ng.g^{-1}). In the DongNai river, the concentrations of indeno[1,2,3-cd]pyrene are highest in SG4 (57.0 ng.g^{-1}) and lowest in DN2 (21.0 ng.g^{-1}). The average indeno[1,2,3-cd]pyrene concentration of all the samples is 33.1 ng.g^{-1}. The indeno[1,2,3-cd]pyrene concentration of the intersection is 28.0 ng.g^{-1}, similar level to samples SG2 & DN1 and lower than the average.

Benzo[a]anthracene+Chrysene :

The benzo[a]anthracene+chrysene concentrations of the sediment samples from 8 locations are from 40.0 to 95.0 ng.g^{-1}. The concentrations of benzo[a]anthracene+chrysene

are highest in SG4 (95.0 ng.g^{-1}) and lowest in DN2 (40.0 ng.g^{-1}). In the SaiGon river, the benzo[a]anthracene+chrysene concentrations are highest in SG3 (75.0 ng.g^{-1}) and lowest in SG1 & SG6 (46.0 ng.g^{-1}). In the DongNai river, the concentration of benzo[a]anthracene+chrysene in SG4 is highest 95.0 (ng.g^{-1}) and lowest in DN2 (40.0 ng.g^{-1}). The average benzo[a]anthracene+chrysene concentration of all the samples is 56.1 ng.g^{-1}. The benzo[a]anthracene+chrysene concentration of the intersection is 49.0 ng.g^{-1}, similar to that in DN1 and lower than the average.

Benzo[b&k]fluoranthene:

The benzo[b&k]fluoranthene concentrations of the sediment samples from 8 locations are from 40.0 to 104.0 ng.g^{-1}. The concentrations of benzo[b&k]fluoranthene are highest in SG4 (104.0 ng.g^{-1}) and lowest in DN2 (40.0 ng.g^{-1}) for all the samples. In the SaiGon river, benzo[b&k]fluoranthene are highest in SG3 (78.0 ng.g^{-1}) and lowest in SG6 (48.0 ng.g^{-1}). In the DongNai river, the benzo[b&k]fluoranthene concentration in SG4 is highest (104.0 ng.g^{-1}) and lowest in DN2 (40.0 ng.g^{-1}). The average benzo[b&k]fluoranthene concentration of all the samples is 61.1 ng.g^{-1}. The benzo[b&k]fluoranthene concentration of the intersection is 53.0 ng.g^{-1}, similar that in SG2 and lower than the average.

Total PAHs :

The total PAHs concentrations of the sediment samples from 8 locations are from 293.0 to 578.0 ng.g^{-1}. The concentration of total PAHs are highest in SG4 (578.0 ng.g^{-1}) and lowest in DN2 (293.0 ng.g^{-1}). In the SaiGon river, the total PAHs in SG3 are highest (540.0 ng.g^{-1}) and lowest in SG6 (337 ng.g^{-1}). In the DongNai river, total PAHs are highest in SG4 (578.0 ng.g^{-1}) and lowest in DN2 (293.0 ng.g^{-1}). The average total PAHs concentration of all the samples is 414.25 ng.g^{-1}. The total PAHs concentration of the intersection is 340.0 ng.g^{-1}, similar to that in SG6 and lower than the average.

Overall:

Contamination of naphthalene, anthracene**,** dibenz(a,h)anthracene and benzo[g,h,i]perylene were close to and above the TEL. The contamination of these PAH compounds were in the concern and should be monitored.

Naphthalene and dibenz(a,h)anthracene concentrations in sample from SG1 location are highest. Anthracene concentrations in the sample from SG2 location are highest. Naphthalene, dibenz(a,h)anthracene and anthracene levels in these sites should be further examined and monitored.

Sample in location SG4 had the highest concentrations of fluoranthene, , benzo[a]pyrene, benzo[g,h,i]perylene, indeno[1,2,3-cd]pyrene, benzo[a]anthracene + chrysene, benzo[b&k]fluoranthene and total PAHs, with pyrene ranked 2nd. Similarly, sample in location SG3 had elevated concentrations of indeno[1,2,3-cd]pyrene, benzo[a]anthracene + chrysene, benzo[b&k] fluoranthene, fluoranthene, benzo[a]pyrene, benzo[g,h,i]perylene and total PAHs, ranked 2nd and with pyrene ranked 1st. The source of contamination of PAHs in these locations should be carefully surveyed.

Table 3.24: 13 PAHs compounds and total PAHs concentration in 8 sediment samples (ng.g^{-1} dry wt) Note: ND = non detectable < 1 ng.g^{-1} dry weight.

Location	PAH ng.g^{-1} dry weight													
	2 rings			3 rings			4 rings	5 rings	6 rings					
	Naphthalene	Acenaphthylene	Acenaphthene	Phenanthrene	Anthracene	Fluoranthene	Pyrene	Benzo[a]pyrene	Dibenz(a,h)anthracene	Benzo[g,h,i]perylene	Indeno[1,2,3-cd]pyrene	Benzo[a]anthracene+Chrysene	Benzo[b&k]fluoranthene	Total PAHs
SG1	133.0	ND	ND	ND	37.0	38.0	28.0	23.0	35.0	31.0	33.0	46.0	57.0	461
SG2	82.0	ND	ND	ND	65.0	47.0	42.0	22.0	23.0	29.0	27.0	50.0	54.0	441
SG3	42.0	ND	ND	ND	57.0	52.0	69.0	33.0	26.0	59.0	49.0	75.0	78.0	540
SG6	46.0	ND	ND	ND	45.0	35.0	31.0	22.0	19.0	22.0	23.0	46.0	48.0	337
DN1	33.0	ND	ND	ND	37.0	28.0	32.0	23.0	17.0	24.0	27.0	48.0	55.0	324
DN2	25.0	ND	ND	ND	52.0	28.0	29.0	20.0	19.0	19.0	21.0	40.0	40.0	293
SG4	26.0	ND	ND	ND	45.0	54.0	67.0	43.0	22.0	65.0	57.0	95.0	104.0	578
SG5	36.0	ND	ND	ND	36.0	30.0	40.0	23.0	18.0	27.0	28.0	49.0	53.0	340
Average	52.9	ND	ND	ND	46.8	39.0	42.3	26.1	22.4	34.5	33.1	56.1	61.1	414.25
TEL	34.6a	5.87	6.71	86.7	46.9	113.0	153.0	88.8	6.22	N	N	N 108	N	N
PEL	391.0	128.0	88.9	544.0	245.0	1494	1398	763.0	135.0	N	N	N 846	N	N

3.2.1.4. PCBs:

Polychlorinated biphenyls (PCBs) are synthetic organic chemicals of chlorine attached to biphenyl, which is a molecule composed of two benzene rings. There are 209 configurations of PCBs with 1 to 10 chlorine atoms. The chemical formula of PCB is $C_{12}H_{10-x}Cl_x$ (10).

The PCBs concentrations of the SG-DN river sediment in 02-2012 ranged from 0.44 $ng.g^{-1}$ to 1.8 $ng.g^{-1}$. PCBs concentrations in the SaiGon river are highest in SG3 (1.7 $ng.g^{-1}$) and lowest is SG1 (0.44 $ng.g^{-1}$). PCBs concentrations in the DongNai river are highest in DN1 (1.8 $ng.g^{-1}$) and lowest in DN2 (0.73 $ng.g^{-1}$). The average concentration of PCBs of the SG-DN river is 1.08 $ng.g^{-1}$. Overall, the concentrations of PCBs in the SG-DN river are very low regarding to the TEL (1.08 $ng.g^{-1}$ / 21.5 $ng.g^{-1}$).

3.2.2. Data obtained on August 2012:

3.2.2.1. PAHs:

Among 17 standard PAH compounds analyzed, acenaphthene, dibenz(a,h)anthracene, benzo(j)fluoranthene and benzo(e)pyrene were not detected with a detection thresold of 1 $ng.g^{-1}$ dry weight. The concentrations of napthaplene (from 64 $ng.g^{-1}$- 225 $ng.g^{-1}$, average 120.4 $ng.g^{-1}$), perylene (from 23 $ng.g^{-1}$-807 $ng.g^{-1}$, average 255.6 $ng.g^{-1}$), anthracene (from 36 $ng.g^{-1}$ -712 $ng.g^{-1}$, average 108.9 $ng.g^{-1}$), fluoranthene (from 6 $ng.g^{-1}$-720 $ng.g^{-1}$, average 57.7 $ng.g^{-1}$) and pyrene (from 5 $ng.g^{-1}$-702 $ng.g^{-1}$, average 64.5 $ng.g^{-1}$) were present for all samples. Several PAH compounds do not appear in 22 samples and in lesser quantities such as benzo[a]anthracene + chrysene (17 out of 22 samples, from 7 $ng.g^{-1}$-454 $ng.g^{-1}$), benzo[b&k]fluoranthene (9 samples, from 7-350 $ng.g^{-1}$), acenaphthylene (4 samples, from 27 $ng.g^{-1}$-42 $ng.g^{-1}$) , benzo[a]pyrene (4 samples, from 7 $ng.g^{-1}$- 237 $ng.g^{-1}$) and fluorene (2 samples, from 6 $ng.g^{-1}$- 54 $ng.g^{-1}$). Phenanthrene, benzo[g,h,i]perylene and indeno[1,2,3-cd]pyrene are present in just one out of 22 samples with concentrations of 19 $ng.g^{-1}$, 122 $ng.g^{-1}$ and 138 $ng.g^{-1}$, respectively. The total PAHs concentration varies among the different sites with a range from 194 $ng.g^{-1}$ to 3854 $ng.g^{-1}$ dry weight. Perylene, anthracene, fluoranthene and pyrene vary significantly among the samples, ranging from 23-807 $ng.g^{-1}$, 36 -712 $ng.g^{-1}$, 6-720 $ng.g^{-1}$ and 5-702 $ng.g^{-1}$, respectively. The sample SG8a1 according to the canal site SG8 has the highest concentration of total PAHs (3854 $ng.g^{-1}$) among the samples (**Table 3.25**).

PCA analysis with different plotting types, which are called GG and CC plots (**357**), were performed with PAHs of 22 sediment samples in order to identify the samples behavior due to PAHs components. Seven PAH compounds, napthaplene, perylene,

anthracene, fluoranthene, pyrene, benzo[a]anthracene + chrysene, benzo[b&k]fluoranthene, and total PAHs were included in the PCA analysis due to their presences in the samples.

The PCA GG plot showed that the grouping of PAHs compounds such as anthracene, fluoranthene, pyrene, benzo [a]pyrene, benzo[g,h,I]perylene, indeno[1,2,3-cd]pyrene, benzo[a]anthracene + chrysene, benzo[b&k]fluoranthene and total PAHs correlate with sample SG8a1. This agreed with the fact that SG8a1 had highest concentration of these PAH compounds and total PAHs (**Table 3.25**). It appears that sample SG9a1 had the highest concentration of naphthalene in the GG plot. In fact, sample DN3a1 had the highest concentration of naphthalene (225 $ng.g^{-1}$) and sample SG9a1 ranked 2^{nd} (198 $ng.g^{-1}$). Sample SG9a1 also had high concentration of other PAHs compounds and total PAHs, ranking 2^{nd} after sample SG8a1 (with the factor of 3 for total PAHs concentration). They include anthracene, fluoranthene, pyrene, benzo [a]pyrene, benzo[a]anthracene + chrysene and benzo[b&k]fluoranthene. Samples SG2a1 & SG2b1 correlate with perylene compound. These samples had the highest concentration of perylene (807 and 630 $ng.g^{-1}$, respectively). Other samples that had high perylene concentrations among others are RF1b1, DN1a1 and DN1b1 (601, 449 and 425 $ng.g^{-1}$, respectively).

The lowest concentration of naphthalene occurred in sample DN3b1 (64 $ng.g^{-1}$). Similarly, the lowest concentration of antharacence were found in samples RF2a1 & SG4a1 (36 $ng.g^{-1}$); so as fluoranthene with RF1a1, RF2a1, RF2b1 and SG4a1 (6-8 $ng.g^{-1}$); pyrene with RF1a1, RF2a1 and RF2b1 (5-8 $ng.g^{-1}$); perylene with RF2a1 and RF2b1 (41 and 23 $ng.g^{-1}$, respectively).

Table 3.25: PAHs (ng.g^{-1} dry weight) analysis of sediment samples for the first sample of left side (a1) and the first sample of right side (b1) of 13 locations. Totally there are 22 sediment samples with 17 PAHs compounds were analyzed (ND with detection threshold of 1 ng.g^{-1}dry weight).

Location		Total PAHs	Naphtha-lene	Acenaph-thylene	Fluoren e	Phenan-threne	Anthr a-cene	Fluoran-thene	Pyrene	Benzo[a] pyrene	Benzo[g,h,I] perylene	Indeno [1,2,3-cd] pyrene	Benzo[a] anthracene +Chrysene	Benzo[b&k] fluoranthene	Perylene
SaiGon branch	RF1a1	415	87	ND	ND	ND	50	6	5	ND	ND	ND	ND	ND	267
	RF1b1	885	130	27	ND	ND	94	20	13	ND	ND	ND	ND	ND	601
	SG1a1	512	95	ND	ND	ND	49	27	26	ND	ND	ND	19	7	289
	SG2a1	1034	78	ND	ND	ND	78	34	22	ND	ND	ND	15	ND	807
	SG2b1	799	71	ND	ND	ND	47	20	19	ND	ND	ND	12	ND	630
	SG3a1	412	70	ND	ND	ND	85	44	61	7	ND	ND	25	27	93
	SG3b1	379	114	ND	ND	ND	45	34	54	ND	ND	ND	ND	ND	132
	SG6a1	458	99	ND	ND	ND	101	26	39	ND	ND	ND	33	22	138
DongNai branch	RF2a1	194	104	ND	ND	ND	36	7	6	ND	ND	ND	ND	ND	41
	RF2b1	228	130	ND	ND	ND	60	7	8	ND	ND	ND	ND	ND	23
	DN1a1	663	84	ND	ND	ND	65	16	23	ND	ND	ND	13	13	449
	DN1b1	719	141	ND	ND	ND	95	17	27	ND	ND	ND	14	ND	425
	DN2a1	750	184	ND	ND	ND	139	49	62	ND	ND	ND	33	27	256
	DN2b1	486	156	ND	ND	ND	86	14	21	ND	ND	ND	8	ND	201
	DN3a1	730	225	ND	ND	19	76	40	65	11	ND	ND	51	31	212
	DN3b1	387	64	ND	ND	ND	50	13	20	ND	ND	ND	7	ND	233
	SG4a1	343	87	ND	ND	ND	36	8	20	ND	ND	ND	14	9	169
	SG4b1	441	131	42	6	ND	73	15	27	ND	ND	ND	18	ND	129
Junction	SG5a1	465	126	42	ND	ND	101	16	24	ND	ND	ND	15	ND	141
	SG5b1	325	97	ND	ND	ND	78	11	15	ND	ND	ND	11	ND	113
Canals	SG8a1	3854	177	29	54	ND	712	720	702	237	122	138	454	350	159
	SG9a1	1093	198	ND	ND	ND	239	126	159	41	ND	ND	127	88	115

In order to investigate whether different components could affect the grouping of the samples according to 7 most dominant PAH components and total PAHs, different PCA analyses in CC plot with PC1 & PC2, PC2 & PC3 and PC1 & PC3 were performed for i) 22 samples with canal samples SG8a1 & SG9a1, ii) 21 samples without sample SG8a1, iii) 20 samples without SG8a1 & SG9a1 (**Fig. 3.14-22**, subsequently).

i) With SG8a1 & SG9a1:

First, PCA analysis of PAHs from 22 sediment samples were performed (**Fig. 3.14-16**). The results showed that canal samples SG8a1 and SG9a1 separated from the rest of samples on PC1 component with the variance 76.3 %. This agreed with the results from PCA with GG plot (**Fig. 3.13**). PC2 that explained 14.5% distinguished samples of the SaiGon river, SG2a1 & SG2b2 and agreeable the results from PCA with GG plot (**Fig. 3.12**). The highest concentration of perylene in samples SG2a1 & SG2b2 explained the separation of these samples (**Table 3.25**).

ii) Without SG8a1:

Since sample SG8a1 had extremely high PAHs concentrations compared to the other samples, it was eliminated in PCA analysis in order to see clearer the correlation among the others. PCA analysis on PAHs of 21 sediment samples were performed without the present of sample SG8a1. Samples of DongNai river, DN2a1 & DN3a1, separate from other samples in PC1 with 67.3% variance (**Fig. 3.17-19**). These samples have similar PAHs concentration, including napthalene, fluoranthene, pyrene, benzo[a]anthracene + chrysene, perylene and total PAHs (225 & 184 $ng.g^{-1}$, 40 & 49 $ng.g^{-1}$, 65 & 62 $ng.g^{-1}$, 51 & 33 $ng.g^{-1}$, 212 & 256 $ng.g^{-1}$ and 730 & 750 $ng.g^{-1}$, respectively). Sample DN3a1 has the highest concentration of napthalene (**Table 3.25**).

iii) Without SG8a1 & SG9a1:

Similarly, canal samples SG8a1& SG9a1 were eliminated in order to reveal the relationship among the river samples. PAHs analysis of 20 sediment samples with PC1 & PC2, PC1 & PC3 and PC2 & PC3 were performed (**Fig. 20-22**).

Results of PCA with PC1 & PC2 of which explained 51.1% & 24.4% variance, samples of the DongNai river, DN2a1 & DN3a1, and samples of the SaiGon river SG3a1 & SG6a1 separate from the rest of samples. Similar to the grouping of samples DN2a1 & DN3a1, samples SG3a1 & SG6a1 had similar concentration of naphthalene, anthracene,benzo[a]anthracene + chrysene, benzo[b&k]fluoranthene, perylene and total PAHs (70 & 99 $ng.g^{-1}$, 85 & 101 $ng.g^{-1}$, 25 & 33 $ng.g^{-1}$, 27 & 22 $ng.g^{-1}$, 93 & 138 $ng.g^{-1}$ and 412 & 458 $ng.g^{-1}$, respectively). Similarly, samples of the upstream DongNai river,

RF2a1 & RF2b1, clustered because the shared the same characteristics of several PAHs compounds, with the lowest concentration of of fluoranthene, pyrene, perylene and total PAHs (7 & 7 $ng.g^{-1}$, 6 & 8 $ng.g^{-1}$, 41 & 23 $ng.g^{-1}$, 194 & 228 $ng.g^{-1}$, respectively) (**Fig. 3.19**). Similar concentrations of fluoranthene, pyrene, benzo[a]anthracene + chrysene, perylene and total PAHs of samples DN1a1 & DN1b1 resulted the clustering of these samples, 16 & 17 $ng.g^{-1}$, 23 & 27 $ng.g^{-1}$, 13 & 14 $ng.g^{-1}$, 449 & 425 $ng.g^{-1}$ and 663 & 719 $ng.g^{-1}$, respectively) (**Fig. 3.20**, **Table 3.25**).

On the PC1 & PC3, which explained 51.1% & 11.7% variance, clearly showed the separation of samples from the SaiGon and the DongNai rivers with RF1a1, SG1a1, SG2 (a1, b1), SG3 (a1, b1), SG6a1 lies in one side of the PC3 axis while RF2 (a1, b1), DN1b1, DN2 (a1, b1), DN3a1, SG4b1 lies at the other side of the PC3 axis. Samples from intersection location, SG5 (a1, b1) lie on the same side as the DongNai river, suggesting their similar characteristic with DongNai river (**Fig. 3.21**).

The samples RF1a1, SG1a1, SG2 (a1, b1), SG3 (a1, b1) and SG6a1 which belongs to the SaiGon river had the higher total PAHs concentrations than those of samples RF2 (a1, b1), DN1b1, DN2 (a1, b1), DN3a1, SG4b1 from the DongNai river. Naphthalene concentrations in the SaiGon cluster are lower than that of the DongNai cluster, explaining their separation on the PCA plot. Similarly, samples SG5 (a1, b1) had higher concentrations of naphthalene compared to the SaiGon river samples, making them group with the DongNai river samples. Fluoranthene and pyrene concentrations are highest in the sample SG3a1, explaining its distance from other samples in the plot. Sample DN1a1, DN3b1 and SG4a1 shared the lowest naphthalene concentrations of the SaiGon river samples, making them group with the SaiGon samples. The SaiGon river seems to be more polluted with PAH compounds than the DongNai river based on value of total PAHs (**Table 3.25**).

In PC2 (24.2%) & PC3 (11.7%), the SaiGon river presents the order SG2→SG1→RF1→SG3 & SG6. Sample SG3a1 is separated from the other samples (**Fig. 3.22**). In the DongNai river, the river presents as DN1→ DN2 → DN3 → RF2. The evolution of the SaiGon river can be explained by the decreasing total PAHs and perylene concentrations from upstream to downstream of the river. The total PAHs concentrations gradually decrease from sample SG2→SG1→RF1→SG3 & SG6 with the total PAHs concentrations of sample SG6a1 being slightly higher than that of RF1a1, SG3a1 and SG3b1. Similarly, the total PAHs and perylene concentrations decreased from upstream to downstream of the DongNai river (from DN1 to DN3) with the exception that sample

DN3a1. The upstream samples RF2 have the lowest concentration of total PAHs and perylene (**Table 3.25**).

i) With SG8a1 & SG9a1:

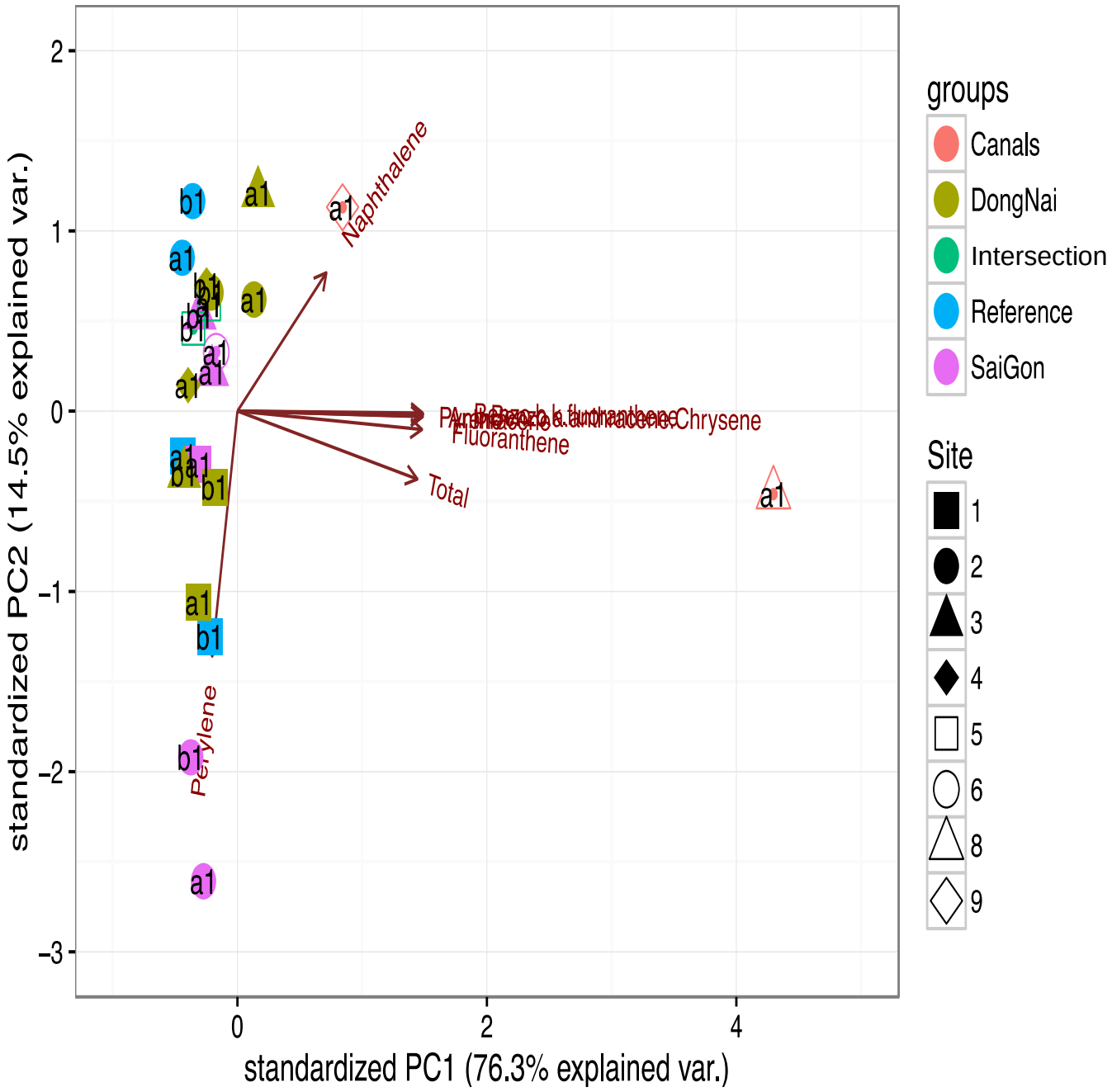


Figure 3.13. PCA GG plot of the chemical analytes (PAHs) of 22 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first two principal components (PC1 & PC2).

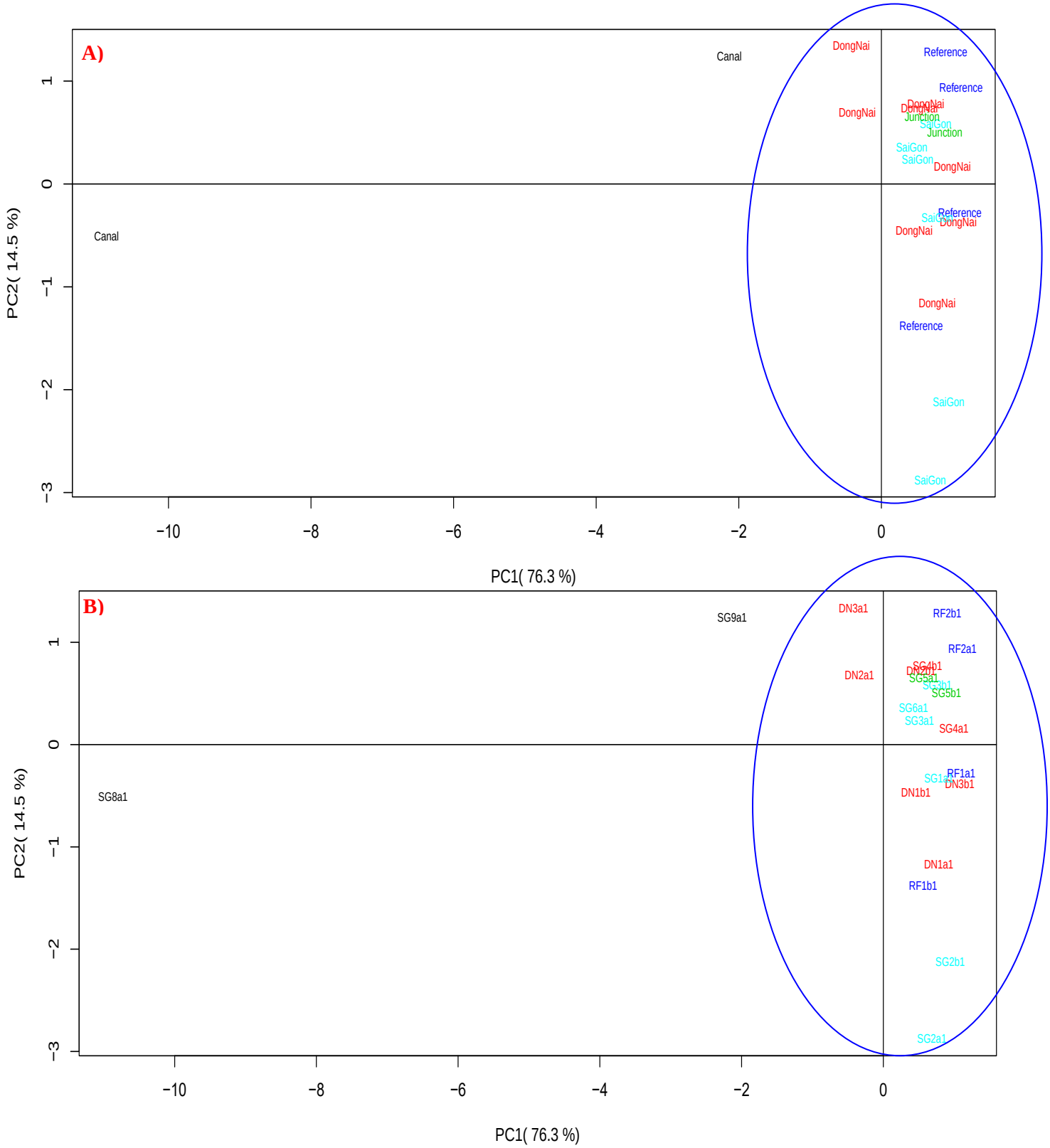


Figure 3.14. PCA CC of plot of the chemical analytes (PAHs) of 22 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first two principal components (PC1 & PC2). Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

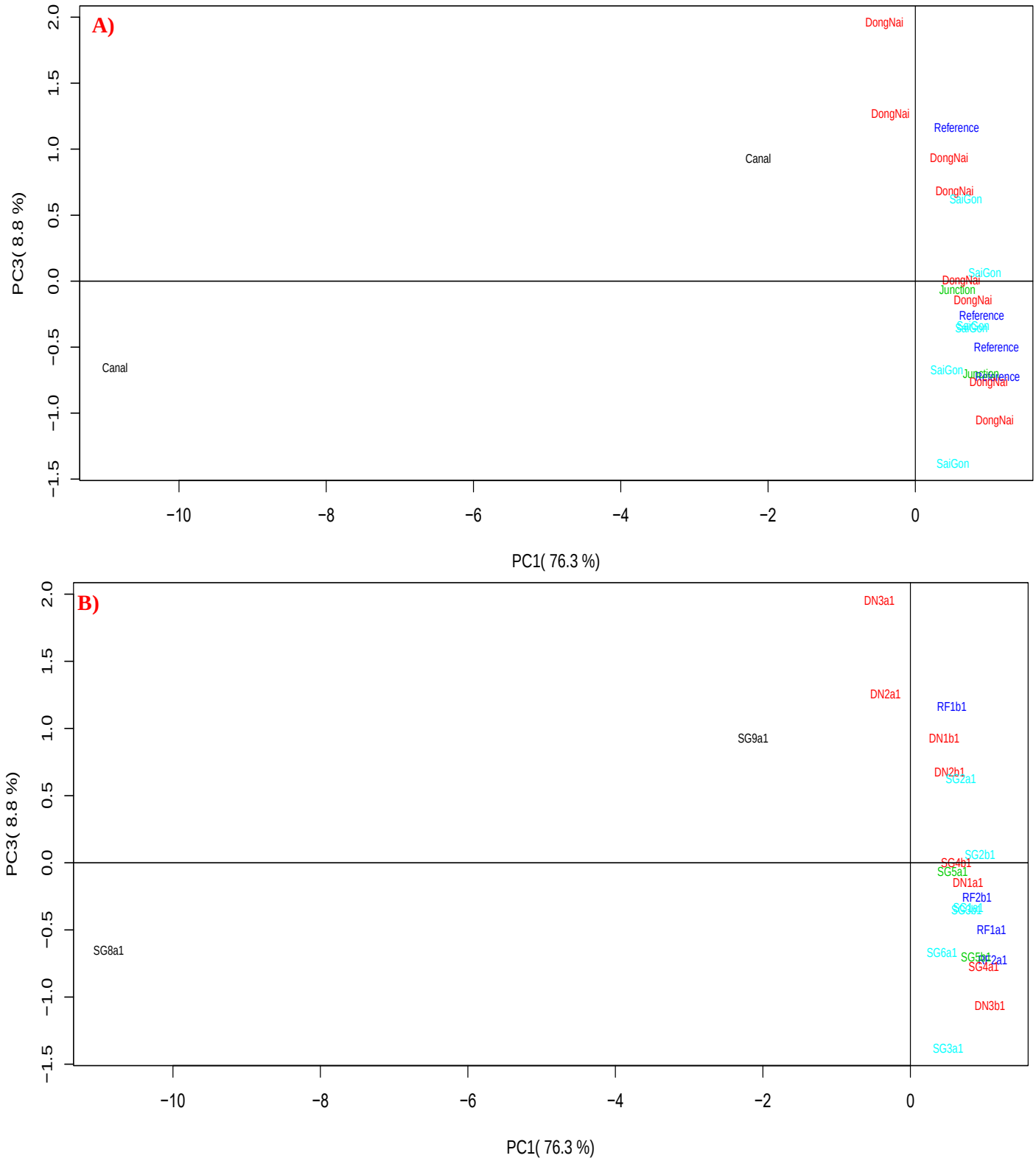


Figure 3.15. PCA CC of plot of the chemical analytes (PAHs) of 22 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first and third principal components (PC1 & PC3). Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

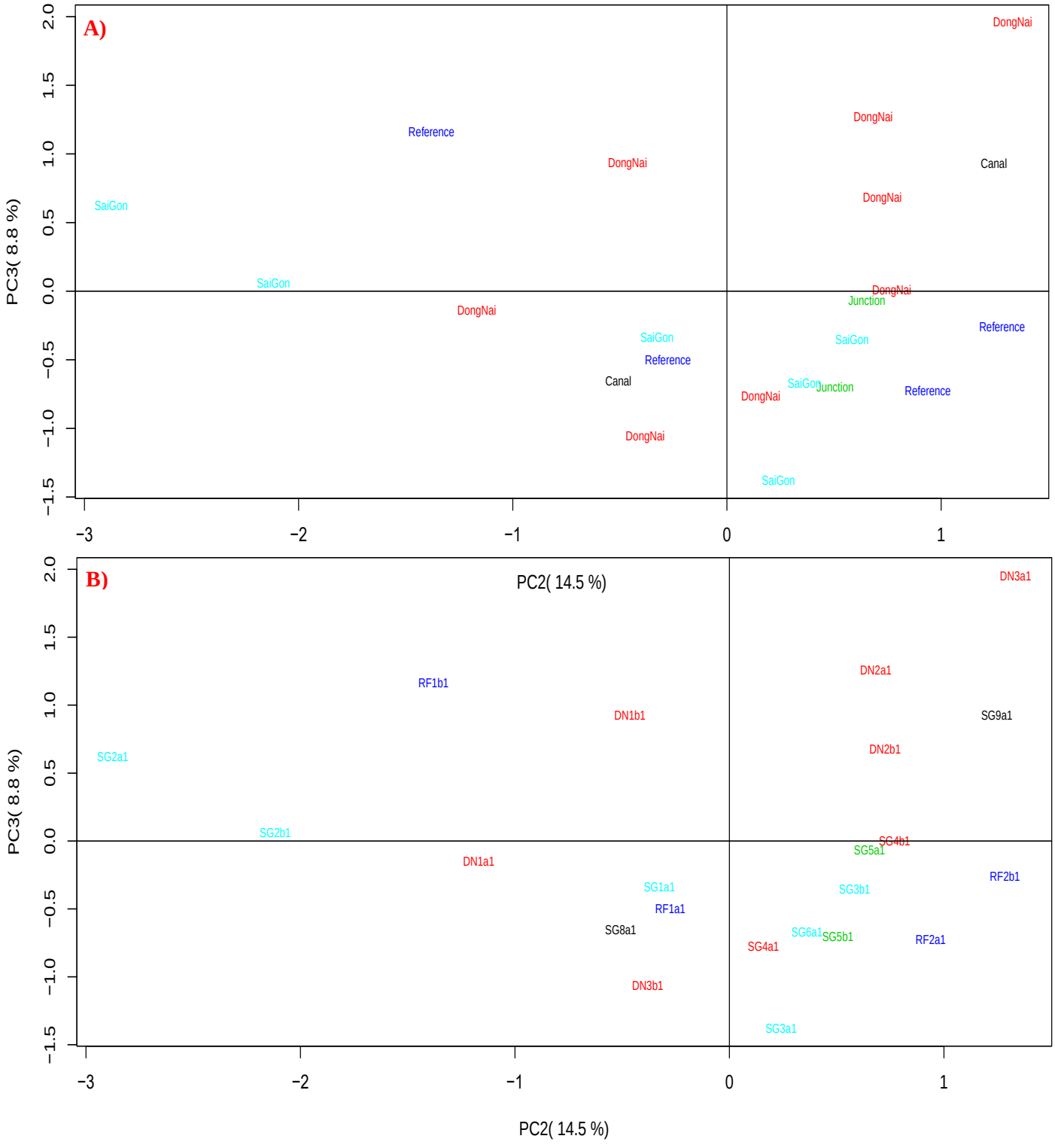


Figure 3.16. PCA CC of plot of the chemical analytes (PAHs) of 22 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the second and third principal components (PC2 & PC3). Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

ii) Without SG8a1:

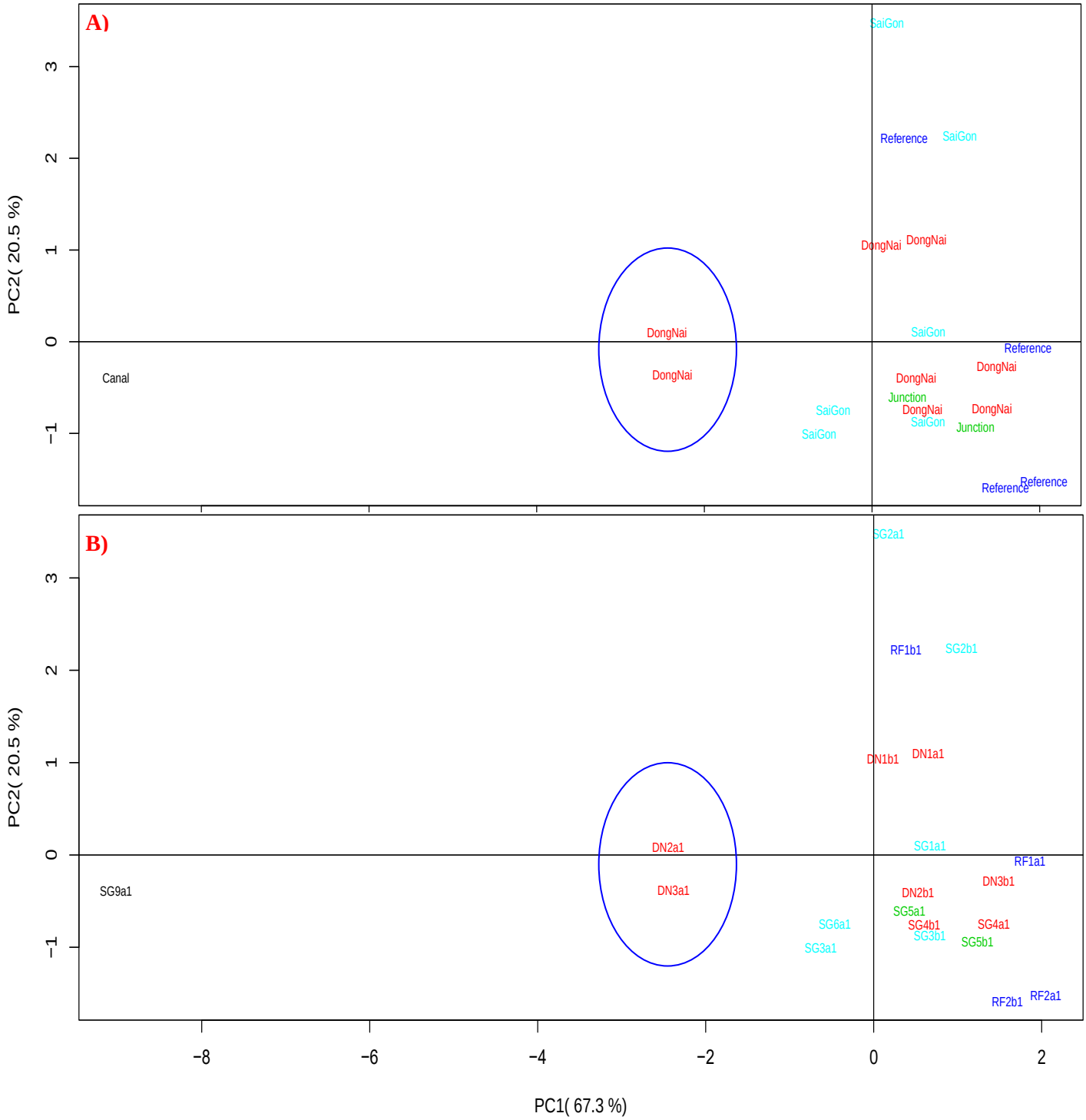


Figure 3.17. PCA CC of plot of the chemical analytes (PAHs) of 21 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first two principal components (PC1 & PC2) with sample SG8a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

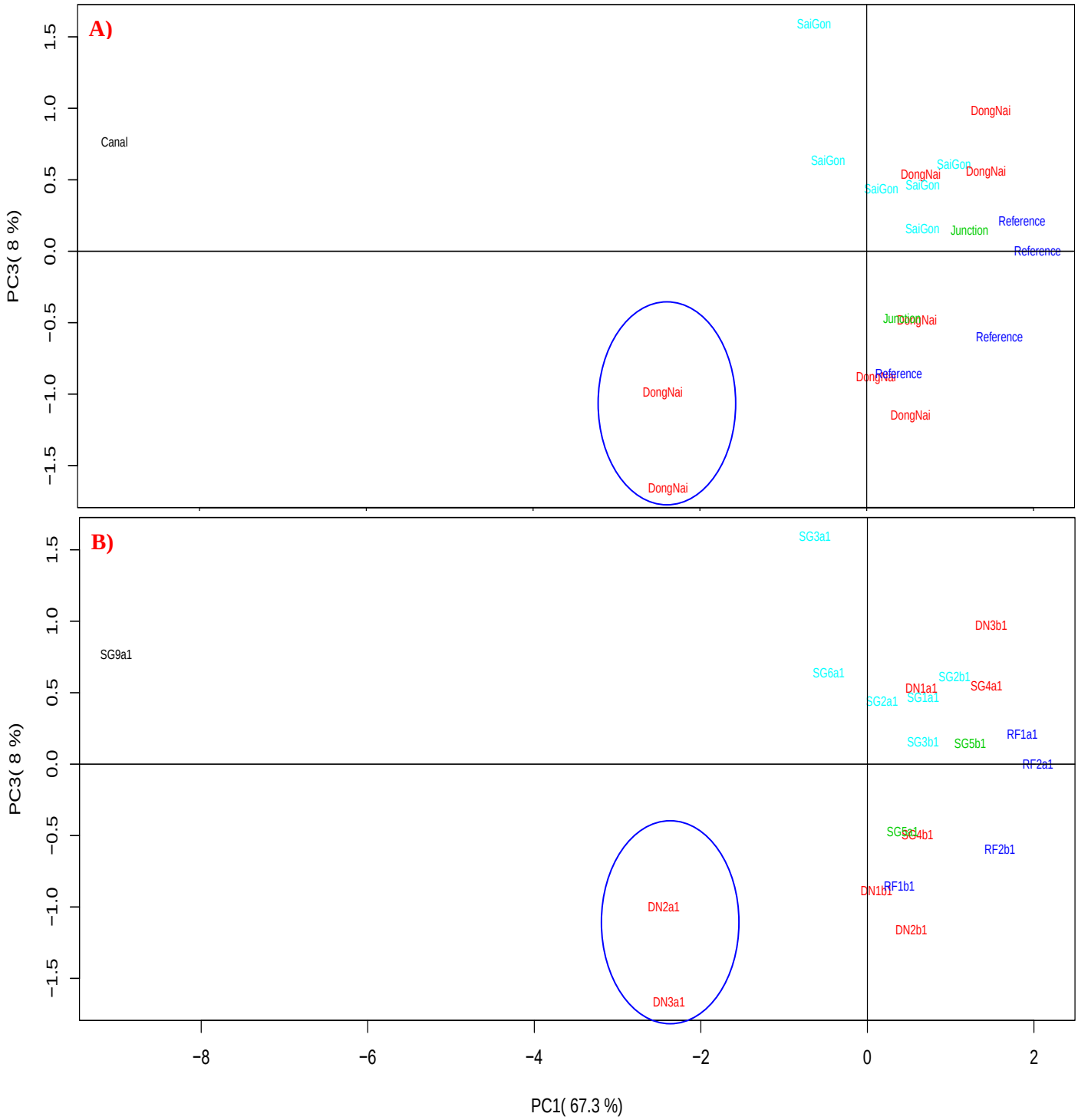


Figure 3.18. PCA CC of plot of the chemical analytes (PAHs) of 21 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first and third principal components (PC1 & PC3) with sample SG8a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

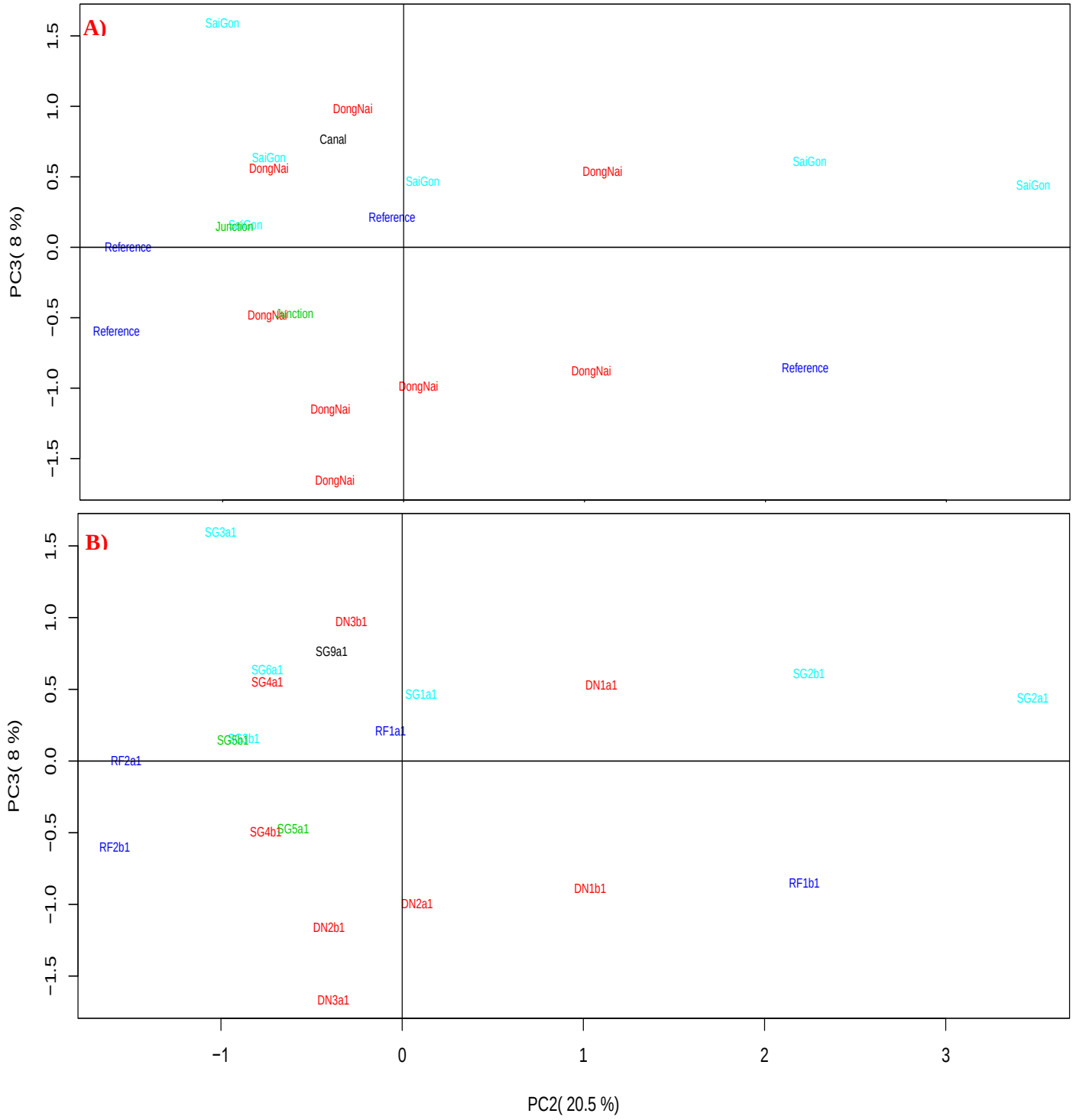


Figure 3.19. PCA CC of plot of the chemical analytes (PAHs) of 21 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the second and third principal components (PC2 & PC3) with sample SG8a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

iii) Without SG8a1 & SG9a1:

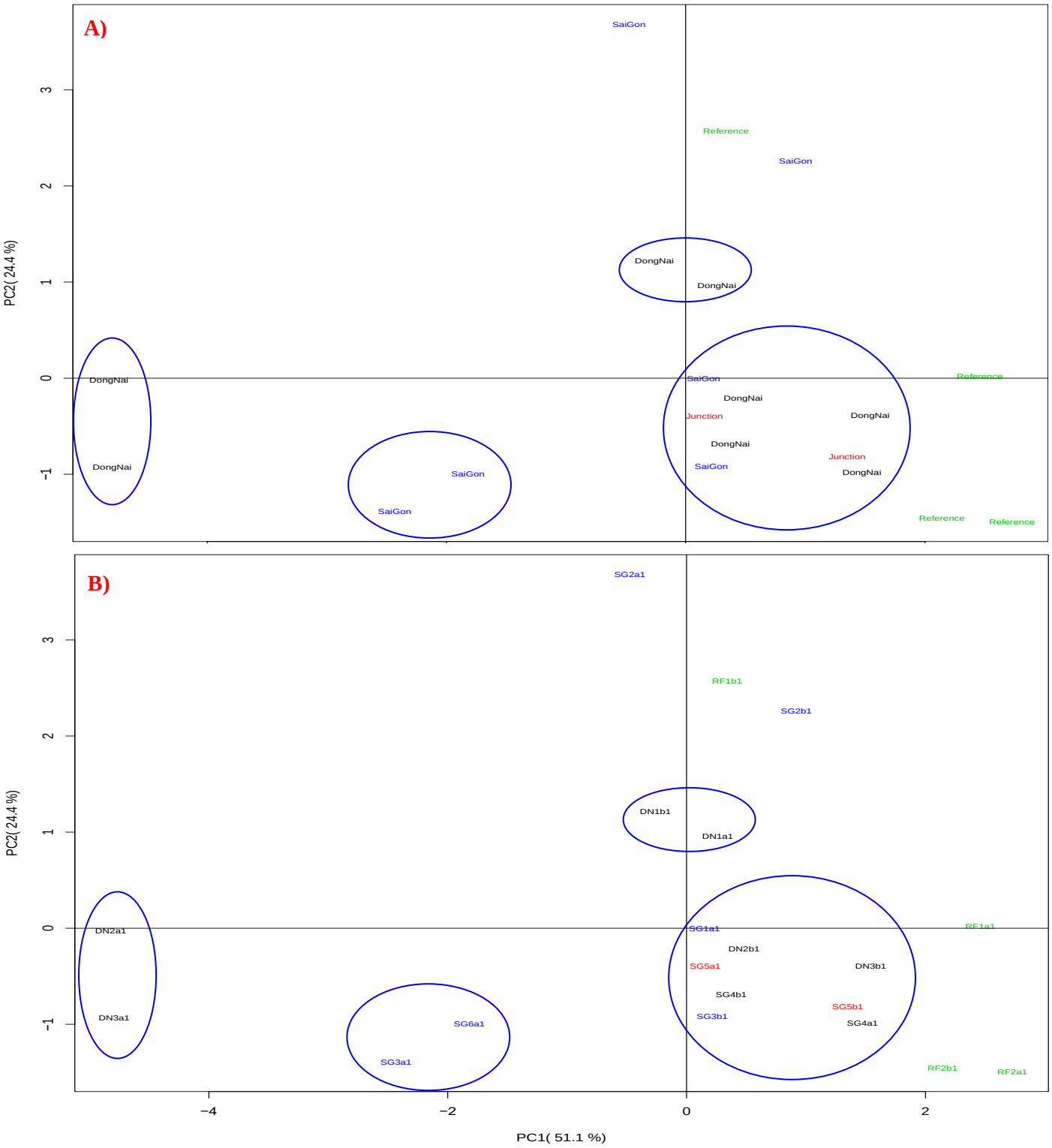


Figure 3.20. PCA CC of plot of the chemical analytes (PAHs) of 20 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first two principal components (PC1 & PC2) with samples SG8a1 & SG9a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

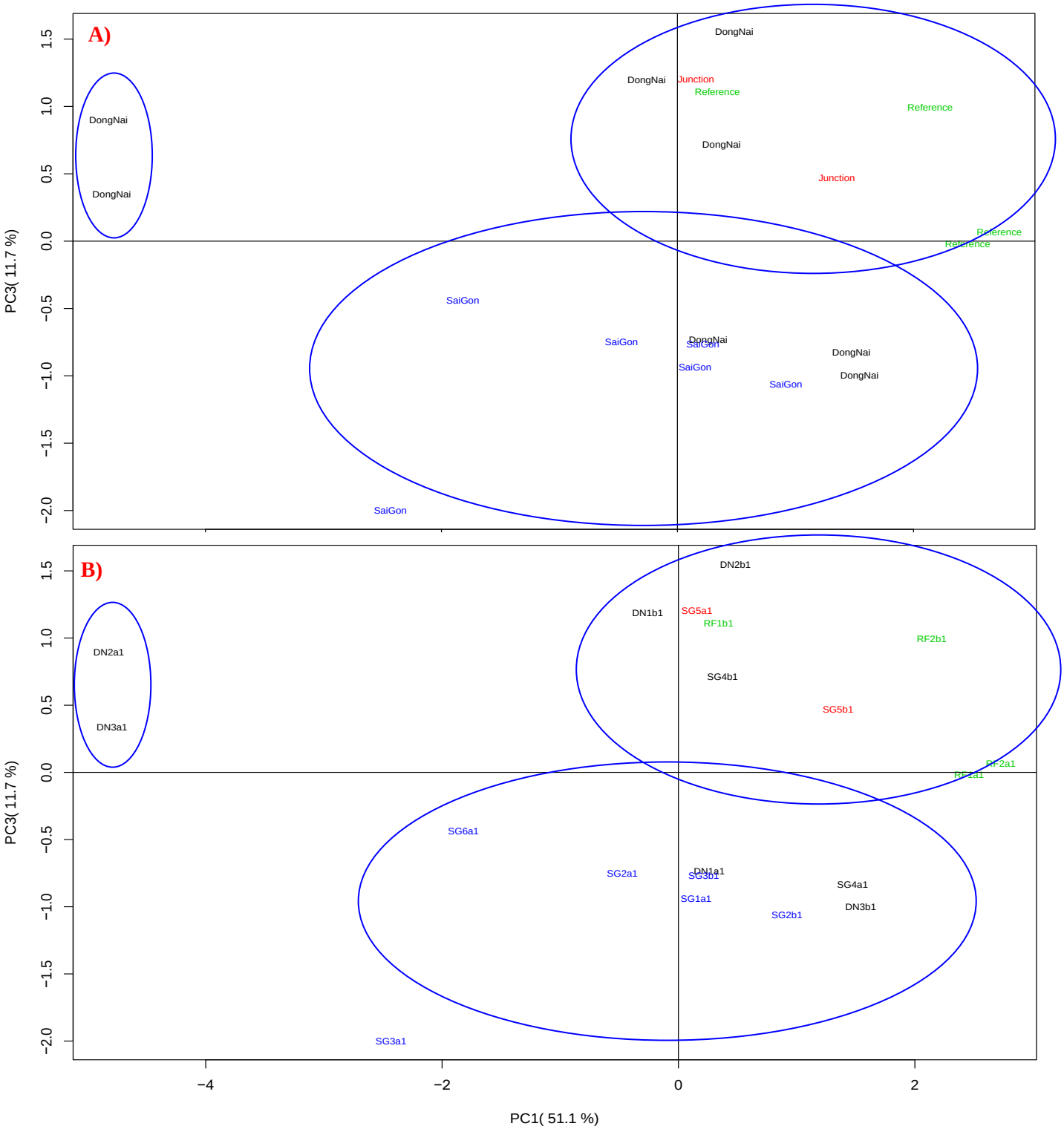


Figure 3.21. PCA CC of plot of the chemical analytes (PAHs) of 20 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the first and third principal components (PC1 & PC3) with samples SG8a1 & SG9a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

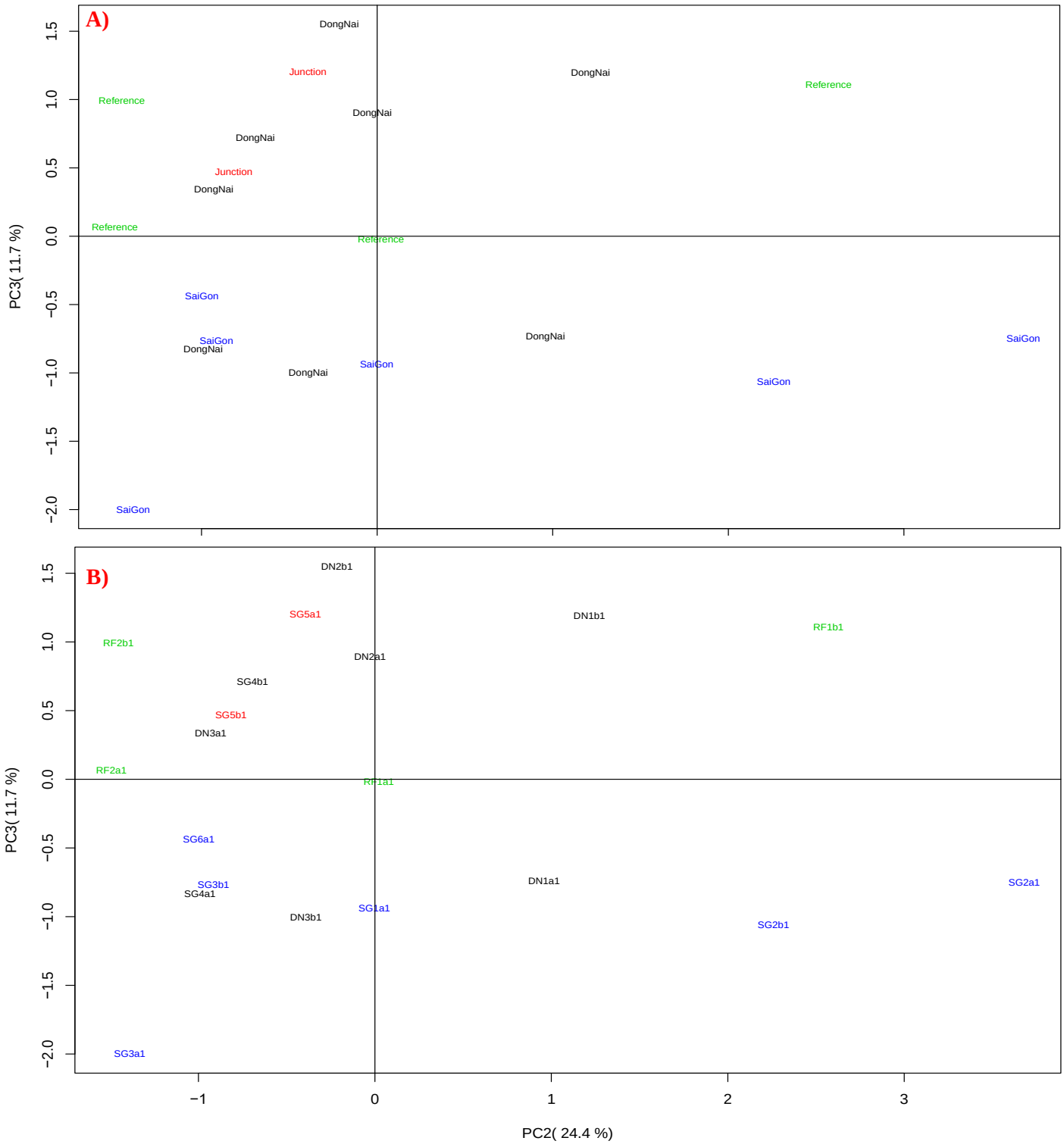


Figure 3.22. PCA CC of plot of the chemical analytes (PAHs) of 20 sediment samples (the first sample of left side (a1) and the first sample of right side (b1) of 13 locations) from the SG-DN river system on the second and third principal components (PC2 & PC3) with samples SG8a1 & SG9a1 removed. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

Overal:

PCA of PAH compounds and total PAHs showed that the two rivers had decreased concentrations from upstream to downstream. With samples SG8a1 & SG9a1 removed, the two rivers clearly seperated in PC1 & PC3 analyses due to their distinct PAHs profiles. Samples from the intersection location SG5 (a1, b1) shared the same characteristics with the DN river the PAH compound levels. The two reference locations which are samples RF1a1 & RF1b1 from the SaiGon river and samples RF2a1 & RF2b1 from the DongNai river the also behaved differently, in which samples RF2a1 & RF2b1 shared more similar concentrations of PAHs than did samples RF1a1 & RF1b1. Samples from canal locations are the most polluted due to their extreme dominant concentration of several PAH compounds and total PAHs.

3.2.2.2. Comparison of PAHs concentrations between 2 sides of the river:

Total PAHs concentrations (17 PAH compounds) of 2 sides of the river: the left side (a1) and the right side (b1) were compared among the samples, except samples from location SG1, SG6, SG8 and SG9 with 1 side of the river being taken due to transportation difficulties (**Fig. 3.23**).

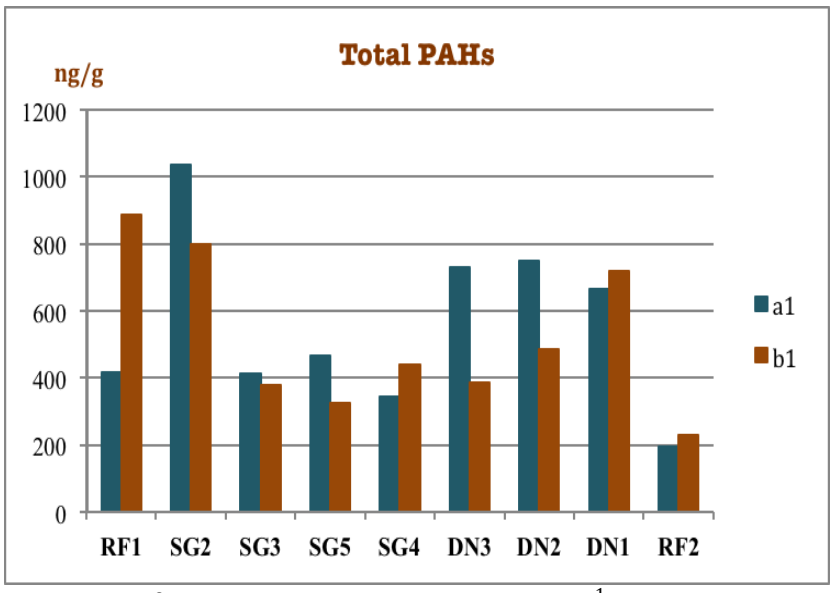


Figure 3.23. Comparison of total PAHs concentrations (ng.g^{-1} dry wt) between the left side (a1) and the right side (b1) of sediment samples from 8 locations in the SG-DN river.

In location RF1, the total PAHs concentration in the sample taken from the b-side is 2.13-fold higher than sample taken from the a-side. Oppositely, in location SG2, the total PAHs concentration of sample SG2a1 is 1.29-fold higher than that of sample SG2b1. In the locations SG3, DN1 and RF2, the total PAHs concentrations are quite similar between the two sides of the river (by 1.08; 0.92 and 0.85-fold). In locations DN2 and

DN3, the total PAH concentrations in the samples taken from the a-side are higher than those from the b-side by 1.54 and 1.88-fold.

Overall, the total PAHs concentrations distributed differently between two sides of the SG-DN river system.

3.2.2.3. Comparison of PAHs concentrations between February 2012 and August 2012 samples:

There are 17 PAH compounds that were analyzed on the samples from August 2012 compared with 13 PAH compounds that were analyzed on the samples taken from February 2012. The four extra PAH compounds are fluorene, benzo(j)flouranthene, benzo(e)pyrene and pyrylene. PAH compounds, including acenapthylene, acenaphthene and phenanathrene were not compared due to their non-detected level in the samples taken on February 2012. The concentrations of each PAH compound and total PAHs were compared among locations at two different time points: February 2012 and August 2012. For this comparison, with the samples taken on August 2012, PAH compounds concentrations were the average of a-side and b-side for each location (**Fig. 3.24**).

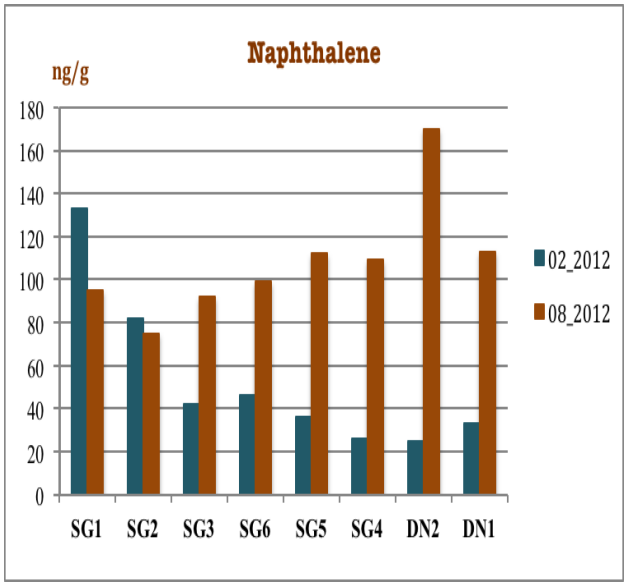
The naphthalene concentrations increased in the samples taken from August 2012 compared to those from February 2012, from 2.2 to 6.8 folds. In contrast, samples of locations SG1 and SG2 were slightly decreased with 1.4 fold and less. The anthracene concentrations increased in the samples taken from August 2012 compared to those from February 2012, from 1.1 to 2.5 fold.

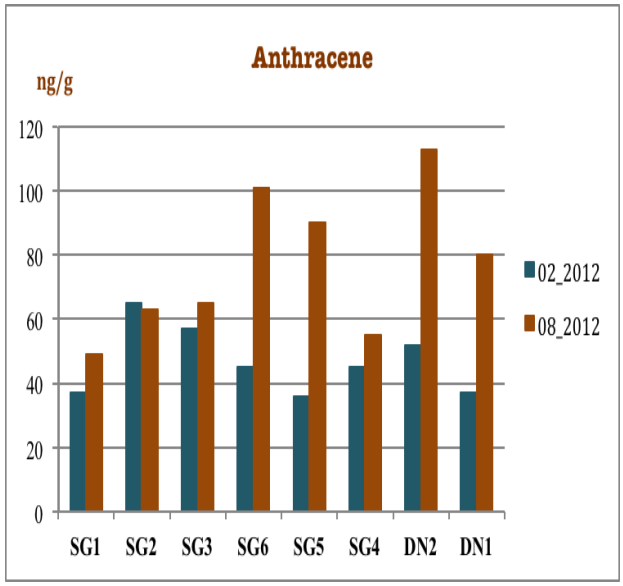
In contrast, the fluoranthene concentrations decreased in the samples taken from August 2012, from 1.3 to 4.5 fold. The fluoranthene concentrations in the samples of location DN2 were slightly increased in samples taken in August 2012.

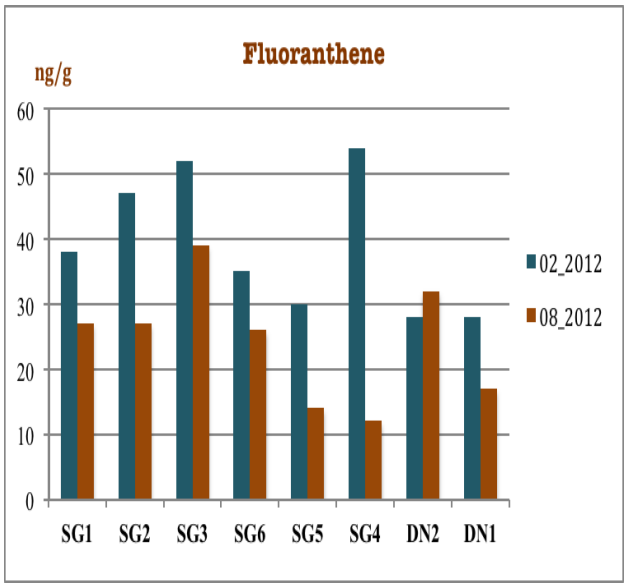
The pyrene concentrations, similar to fluoranthene, decreased in the samples taken from August 2012, from 1.1 to 2.8 fold in the samples of SG1 and SG4 locations, respectively. However, the pyrene concentrations in samples of locations SG6 and DN2 slightly increased.

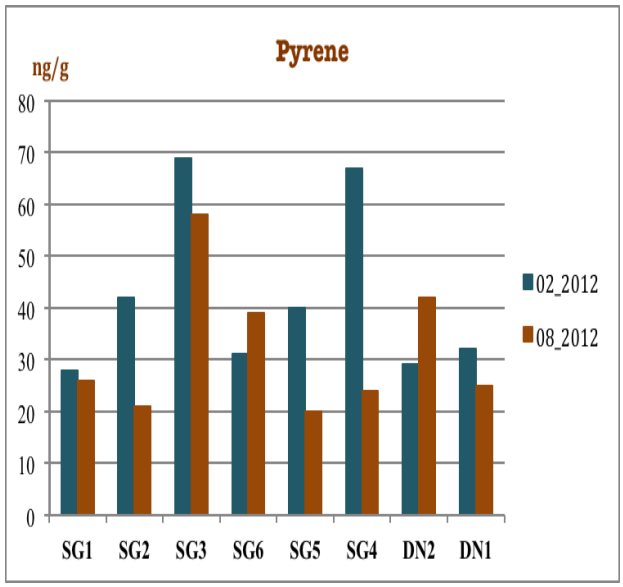
The benzo[a]anthracene + Chrysene concentrations decreased in all the samples taken from August 2012, from 1.4 to 6.0 fold (in samples of SG6 and SG4, respectively).

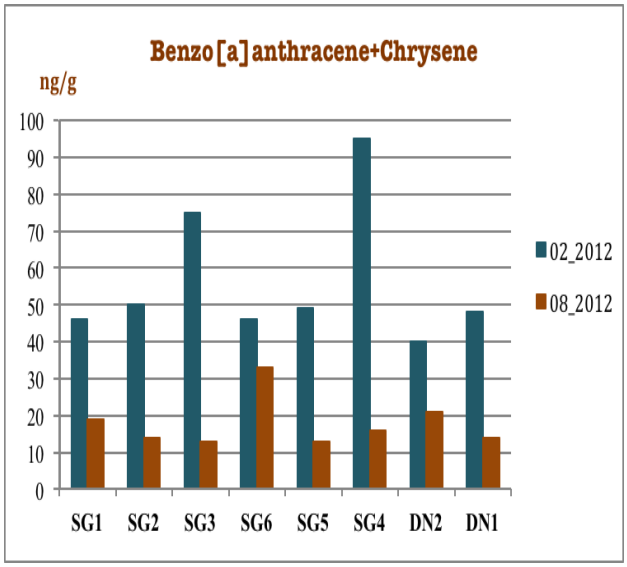
The benzo[b&k]fluoranthene concentrations in all the samples taken from August 2012 are very low compared with samples taken from February 2012 (from 0 $ng.g^{-1}$ - 22 $ng.g^{-1}$ compared with 40 -104 $ng.g^{-1}$, respectively). The decreasing were from 2.2 to 20.8 folds (in samples SG6 and SG4 respectively).

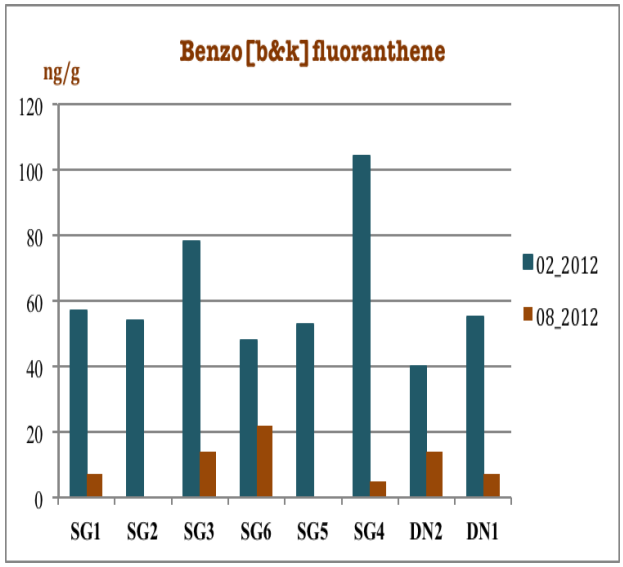

Naphthalene
ng/g
180
160
140
120
100
80
60
40
20
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012


Anthracene
ng/g
120
100
80
60
40
20
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012


Fluoranthene
ng/g
60
50
40
30
20
10
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012


Pyrene
ng/g
80
70
60
50
40
30
20
10
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012


Benzo[a]anthracene+Chrysene
ng/g
100
90
80
70
60
50
40
30
20
10
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012


Benzo[b&k]fluoranthene
ng/g
120
100
80
60
40
20
0
SG1 SG2 SG3 SG6 SG5 SG4 DN2 DN1
02_2012
08_2012

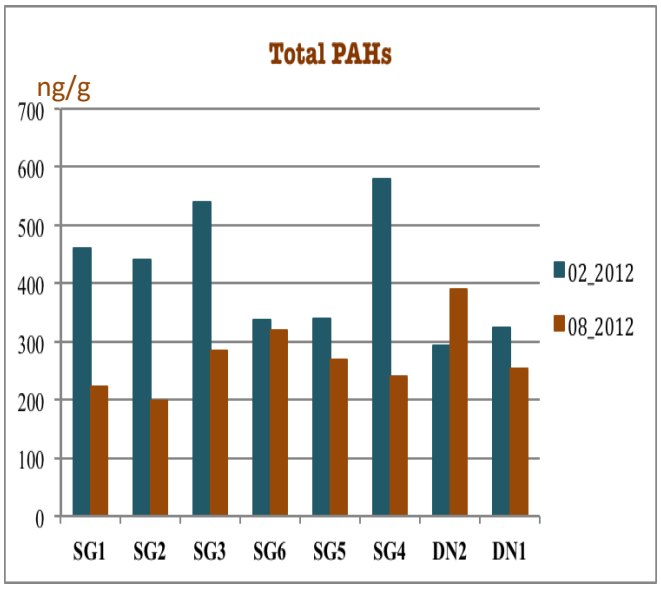


Figure 3.24. Comparison of each PAH compound and total PAHs concentrations (ng.g^{-1} dry wt) between the samples taken on February 2012 and August 2012 in the SG-DN river system.

For the total PAHs concentrations, which are the sum of 13 PAHs compounds, the samples taken from August 2012 decreased compared to February 2012, from 1.1 to 2.4 folds (in samples of SG6 and SG4, respectively). However, total PAHs concentrations in samples of location DN2 increased slightly with 1.3 fold from February to August 2012 (**Fig. 3.24**, **Table 3.29**).

Overal, except for naphthalene and anthracene, the concentration of PAH compounds including fluoranthene, pyrene, benzo[a]anthracene + chrysene, benzo[b&k]fluoranthene and total PAHs decreased in the samples taken from August 2012 compared to February 2012. Naphthalene and anthracene concentrations in DN2 were highest in August 2012 among the samples (**Table 3.29**). Fluoranthene concentrations were highest in SG4 in August 2012 among the samples. Naphthalene and anthracene contamination should be further examined in the SG-DN river system.

Table 3.26: Concentrations of 13 PAH compounds and total PAHs of the samples that were taken from February 2012 and August 2012. Note: for samples that were taken in August 2012, PAH compounds and total PAHs concentrations are the average of the a & b sides of the rivers.

Location	Naphthalene		Anthracene		Fluoranthene		Pyrene		Benzo[a]anthracene + Chrysene		Benzo[b&k] fluoranthene		Total PAHs	
	02-2012	08-2012	02-2012	08-2012	02-2012	08-2012	02-2012	08-2012	02-2012	08-2012	02-2012	08-2012	02-2012	08-2012
SG1	133	95	37	49	38	27	28	26	46	19	57	7	461	223
SG2	82	75	65	63	47	27	42	21	50	14	54	0	441	198
SG3	42	92	57	65	52	39	69	58	75	13	78	14	540	283
SG6	46	99	45	101	35	26	31	39	46	33	48	22	337	320
SG5	36	112	36	90	30	14	40	20	49	13	53	0	340	268
SG4	26	109	45	55	54	12	67	24	95	16	104	5	578	240
DN2	25	170	52	113	28	32	29	42	40	21	40	14	293	390
DN1	33	113	37	80	28	17	32	25	48	14	55	7	324	254

3.2.2.4. Fecal Coliforms & E.coli within the river sediments (data obtained in August 2012):

Fecal coliform values also vary significantly among the samples ranging from 1 to 35,000 MPN.g^{-1} wet weight. Sample DN3b1 belong to the DongNai branch has the highest *Fecal coliform* value 35,000 MPN. g^{-1} wet weight. The lowest concentration of *Fecal coliform* belongs to the samples RF1b1, SG2b1, SG5a1, SG4a1 and DN3a1 (**Table 3.27**).

Samples from the DongNai branch have *Fecal coliform* concentrations higher than those of the SaiGon river. Most of the samples from the DongNai river had *Fecal coliform* concentrations from 290-35000 MPN.g^{-1} wet weight, while in the SaiGon river branch, the concentrations of the majority samples ranked from 30-430 MPN.g^{-1} wet weight (**Table 3.27**).

In the SaiGon river, the *Fecal coliform* concentration are highest in samples SG3a1, then in SG1a1 and SG3b1 (430, 300 and 210 MPN. g^{-1} wet weight, respectively). In the DongNai river, the *Fecal coliform* concentration is highest in samples DN2a1 (35,000 MNP g^{-1} wet weight). The *Fecal coliform* concentrations of the DongNai river increased from SG4→DN1→RF2→DN3→DN2. Samples from the intersection location, SG5 (a1, b1), have very low concentrations of *Fecal coliform* (1-75 MPN. g^{-1}). The canal locations, including SG8a1 & SG9a1 had intermediate concentration of *Fecal coliform* (750 MPN g^{-1} wet weight) (**Table 3.27**).

Table 3.27: *Fecal coliform* analysis of sediment samples for the first sample of left side (a1) and the first sample of right side (b1) of 13 locations. In total, there are 22 sediment samples that were analyzed by the method described in Chapter 2 (Materials & Methods).

Location		*Fecal coliforms* (MPN/g wet weight)	*E. coli* (MPN/g wet weight)
SaiGon branch	RF1a1	30	9
	RF1b1	1	1
	SG1a1	300	1
	SG2a1	30	20
	SG2b1	1	1
	SG3a1	430	210
	SG3b1	210	210
	SG6a1	75	75
DongNai branch	RF2a1	3600	1200
	RF2b1	7000	7000
	DN1a1	350	350
	DN1b1	530	260
	DN2a1	35000	2900
	DN2b1	750	360
	DN3a1	1	1
	DN3b1	19000	19000
	SG4a1	4	4
	SG4b1	290	290
Intersection	SG5a1	1	1
	SG5b1	75	75
Canals	SG8a1	750	750
	SG9a1	750	190

3.3. Pyrosequencing data obtained in August 2012:

Total DNA was extracted from all 42 sediment samples, followed by pyrosequencing of PCR amplified V1-V3 regions of the 16S rRNA gene. A total of 358,392 raw sequences from 3 pyrosequencing runs were obtained for all the samples using the GS Junior system. More than 98% of the sequences were saved after the initial trim (**Table 3.28**). After quality control trimming, a total of 259,983 sequences remained, with saved sequences ranging from 79.87% to 94.64% (**Table 3.29**). Sequences then went through chimeric detection using two different programs (**Materials and Methods**). After removing chimeric sequences, from 66.76% to 88.11% of the sequences were remained. After that, non-bacterial sequences (Archaea and Mitochondria) detected by the Silva database release 123 were also removed, representing < 1 % of the total sequences, with the exception of sample SG9a1 which contained ~ 6% of these types of sequences. After all the cleaning steps, from 66.51% to 88.01% of sequences were saved, with numbers of sequences ranged from 1737 to 10036 (**Table 3.29**).

Table 3.28: Number of raw sequences in 3 runs before and after the initial trim step.

Run	N^0 of raw seqs	Initial trim	
		N^0 of seqs	% saved seqs
I	136776	135315	98.93
II	102826	101675	98.88
III	118790	117372	98.81

Note:

42 PCR products according to 42 sediment samples were and divided into 3 runs and sent to 454-pyrosequencing (GS Junior Sequencing System, University of Oulu, Finland).

After the sequencing, the raw reads from 3 runs (marked as I, II, III) were trimmed from step 1 to 3 (Fig. 2.4, Materials & Methods) and called as Initial Trim process.

Table 3.29: Percentage and number of saved sequences for the 42 sediment samples after each cleaning process.

Location		1. Initial trim	2. Quality Score trim		3. Chimera removal		4. Contaminants removal	
		N^0 of seqs	% of saved seqs	N^0 of saved seqs	% of saved seqs	N^0 of saved seqs	% of saved seqs	N^0 of saved seqs
SaiGon river	RF1_a1	7553	88.73	6702	85.42	6452	85.34	6446
	RF1_a2	10926	90.30	9866	76.60	8369	76.56	8365
	RF1_b1	3174	87.52	2778	71.74	2277	71.74	2277
	RF1_b2	3901	89.90	3507	76.52	2985	76.34	2983
	SG1_a1	9976	90.25	9003	78.33	7814	78.19	7800
	SG1_a2	11256	90.65	10204	75.18	8462	75.10	8453
	SG2_a1	6966	88.69	6178	81.57	5682	81.51	5678
	SG2_a2	2136	87.55	1870	81.46	1740	81.32	1737
	SG2_a3	9059	89.34	8093	84.45	7650	84.42	7648
	SG2_b1	5964	89.02	5309	79.41	4736	79.29	4729
	SG2_b2	9536	90.53	8633	82.61	7878	82.51	7868
	SG2_b3	8663	94.89	8220	88.11	7633	88.01	7624
	SG3_a1	6207	89.33	5545	68.94	4279	68.16	4231
	SG3_b1	7636	90.58	6917	69.21	5285	68.99	5268
	SG6_a1	9494	86.80	8241	67.03	6364	66.48	6312
	SG6_a2	6648	87.17	5795	68.70	4567	68.40	4547
DongNai river	RF2_a1	8751	95.49	8356	72.62	6355	72.56	6350
	RF2_a2	6162	94.76	5839	71.28	4392	71.19	4387
	RF2_b1	13383	93.51	12515	72.77	9739	72.64	9722
	RF2_b2	5975	95.60	5712	72.40	4326	72.33	4322
	DN1_a1	9686	94.63	9166	82.52	7993	82.38	7979
	DN1_a2	12406	95.12	11801	79.22	9828	79.11	9815
	DN1_b1	9136	94.64	8646	84.41	7712	84.26	7698
	DN2_a1	6797	93.00	6321	74.27	5048	74.12	5038
	DN2_a2	8425	94.71	7979	76.96	6484	76.82	6472
	DN2_a3	13172	94.65	12467	76.31	10051	76.19	10036
	DN2_b1	10557	95.13	10043	79.29	8371	79.22	8363
	DN2_b2	9526	95.49	9096	70.51	6717	70.46	6712
	DN2_b3	9151	95.25	8716	76.24	6977	76.19	6972
	DN3_a1	8980	84.83	7618	78.33	7034	78.24	7026
	DN3_a2	9543	85.26	8136	81.75	7801	81.71	7798
	DN3_b1	6125	82.55	5056	68.13	4173	68.08	4170
	SG4_a1	7918	85.17	6744	70.17	5556	70.11	5551
	SG4_b1	8349	85.24	7117	71.52	5971	71.46	5966
	SG4_b2	9972	86.74	8650	71.31	7111	71.28	7108
	SG4_b3	8366	84.71	7087	72.33	6051	72.28	6047
Intersection	SG5_a1	5091	79.87	4066	66.76	3399	66.73	3397
	SG5_a2	9706	85.56	8304	69.38	6734	69.33	6729
	SG5_b1	6702	84.96	5694	68.96	4622	68.76	4608
	SG5_b2	8381	86.24	7228	69.06	5788	68.76	5763
Canals	SG8_a1	5135	89.27	4584	72.37	3716	71.47	3670
	SG9_a1	9507	83.56	7944	72.57	6899	66.51	6323

3.4. Bacterial community on the SaiGon-DongNai river system:

3.4.1. Diversity and richness of 40 sediment samples of the SG-DN river system:

After all the cleaning steps, the number of sequences were then normalized to 2983 for each sample, with the samples SG2a2 and RF1b1 being removed due to their low sequence number < 2983 (**Table 3.29**). Operational Taxonomic Units (OTUs), Chao1 and Shannon indices were calculated on the normalized sequences.

3.4.1.1. OTUs:

The OTU numbers were calculated at 97% similarity levels for 40 samples **(344).**

3.4.1.1.1. Before sequences normalization:

Before sequence normalization, the OTU numbers of the SaiGon river ranged from 2030 to 4736, while those of the DongNai river ranged from 2602 to 6162, showing that the OTU numbers of the samples belonging to the DongNai river were higher than those belonging to the SaiGon river.

In the SaiGon river, the highest diversity was detected in the samples RF1a2, SG1(a1, a2) SG2b2 and SG2b3 with OTU numbers ranging from 4441 to 4736, while the lowest diversity was found in the samples RF1b2, SG2b1, SG3(a1,b1) and SG6a2 with OTU numbers ranging from 2030 to 2862. The OTUs evolution of the SaiGon river was from SG1→ SG2 → SG6 → SG3, with the decreasing number of OTUs from 4736 to 2299. Samples from the upstream location RF1, had the OTU numbers vary significantly from 2030 to 4552, with sample RF1b2 is lowest. Similarly, samples from SG6 location, SG6a1 had more 1000 OTUs than SG6a2 (**Table 3.30**). Except for the behavior of OTU numbers from locations RF1 and SG6, the OTU numbers in the SaiGon river tend to decrease from upstream to downstream.

In the DongNai river, samples RF2b1, DN1a2 and DN2a3 had the highest OTUs numbers among the samples with the range from 5946 to 6162, indicating the high diversity of bacterial communities in these sites. In contrast, samples DN3b1 and RF2b2 had the lowest OTUs numbers (2602 and 2925, respectively). The OTU numbers were high in locations DN2, DN1 and RF2. However, the OTU numbers decreased further downstream, including samples of SG4 and DN3 locations (SG4b2, DN3a1, DN3a2, SG4b3, SG4b1, SG4a1, DN3b1, with the OTU numbers 4265, 4074, 3604, 3510, 3471, 3280 and 2602, respectively. Generally, in both rivers, the OTU numbers tend to decrease as one goes downstream.

In the intersection location, the OTU numbers varied from 2223 to 4198, with sample SG5a1 had the lowest OTU numbers compared to others in this location. Similarly, the OTU numbers did not distribute equally among samples in this location, in which samples SG5a2 & SG5b2 had higher OTU numbers than did the samples SG5a1 & SG5b1.

The OTU numbers in canal samples, which are SG8a1 & SG9a1, were 1773 and 3148, respectively. Interestingly, sample SG8a1 had the lowest OTU numbers among 40 samples.

3.4.1.1.2. After sequences normalization:

The OTU numbers decreased significantly after the normalization to between 1529 and 2429 OTU per sample.

In the SaiGon river, the OTU numbers of samples SG3a1 & SG3b1 are lowest (1761 & 1727, respectively). In the DongNai river, the OTU numbers are low in samples belonging to downstream locations, including SG4b2, DN3a1, SG4b3, SG4b1, DN3b1, SG4a1 and DN3a2. These results are in agreement with those from before normalization, indicating decreasing OTU numbers at downstream of the DongNai river.

The numbers of OTU in the DongNai river were slightly higher than those of the SaiGon river after normalization in the samples that belong to locations RF2, DN1 and DN2. The OTUs numbers of the SaiGon river ranged from 1727 to 2137, while those of the DongNai river ranges from 1799 to 2429.

Samples from location SG5 had the OTU numbers from 1986 to 2274 after normalization, with OTU number of the sample SG5a1 lower than others (1986). This pattern is similar to the results before normalization.

Sample from canal locations, SG8a1 had the numbers of OTU lowest among 40 samples.

3.4.1.1.3. In summary:

Although the of the OTU numbers in 40 the samples vary significantly before and after normalization, the behavior of the OTU numbers remains similar, with lower numbers at the downstream of both rivers, lowest numbers in SG5a1, compared with other samples from its location, higher OTUs number in the DN branch compared to the SG branch, and lowest numbers of OTUs in samples SG8a1.

Table 3.30: Number of OTUs, bacterial richness Chao1 and bacterial diversity Shannon index for the 40 sediment samples.

Location		Raw data			Normalized data**		
	Samples	OTUs*	Chao1	Shannon	OTUs	Chao1	Shannon
SaiGon river	RF1_a1	3154	7466	7.55	1827	5524	7.17
	RF1_a2	4552	12166	7.95	2102	8349	7.40
	RF1_b2	2030	5243	7.40	2030	5243	7.40
	SG1_a1	4584	12161	8.03	2125	7768	7.42
	SG1_a2	4736	12697	8.01	2114	8391	7.41
	SG2_a1	3011	8183	7.40	1849	6226	7.06
	SG2_a3	3666	9534	7.46	1799	6630	7.01
	SG2_b1	2759	7342	7.46	1938	6568	7.21
	SG2_b2	4491	11471	7.98	2117	7838	7.42
	SG2_b3	4441	14257	7.95	2137	9165	7.42
	SG3_a1	2299	6688	7.14	1761	6205	6.97
	SG3_b1	2632	7220	7.18	1727	5897	6.92
	SG6_a1	3889	12817	7.78	2137	9783	7.36
	SG6_a2	2862	9419	7.56	2033	7836	7.30
DongNai river	RF2_a1	4394	13931	8.12	2370	10559	7.62
	RF2_a2	3108	10611	7.77	2240	9251	7.48
	RF2_b1	5946	18249	8.30	2297	10760	7.56
	RF2_b2	2925	9628	7.71	2193	8773	7.48
	DN1_a1	5110	15447	8.21	2301	10666	7.57
	DN1_a2	5952	18452	8.28	2249	11056	7.51
	DN1_b1	5014	15090	8.24	2341	10437	7.61
	DN2_a1	3291	10683	7.78	2138	8225	7.43
	DN2_a2	4096	12830	7.95	2192	9320	7.46
	DN2_a3	6162	19641	8.31	2283	11045	7.53
	DN2_b1	5670	18170	8.39	2429	11848	7.68
	DN2_b2	4392	14976	7.97	2225	11062	7.41
	DN2_b3	4690	15613	8.20	2354	10416	7.63
	DN3_a1	4074	11502	7.89	2101	8665	7.38
	DN3_a2	3604	9435	7.55	1799	5948	7.11
	DN3_b1	2602	9545	7.25	1977	7787	7.04
	SG4_a1	3280	11005	7.39	1966	8312	7.05
	SG4_b1	3471	11660	7.51	1981	8653	7.12
	SG4_b2	4265	14653	7.74	2136	10015	7.25
	SG4_b3	3510	10774	7.56	1984	7607	7.14
Junction	SG5_a1	2223	8227	7.20	1986	7451	7.12
	SG5_a2	4198	15177	7.86	2172	10683	7.31
	SG5_b1	3245	11710	7.84	2274	9467	7.55
	SG5_b2	3862	13245	7.95	2238	10698	7.50
Canals	SG8_a1	1773	4483	6.80	1529	4157	6.74
	SG9_a1	3148	9476	6.88	1679	7930	6.57

*Abbreviation, OTU (Operational Taxonomic Unit)

**The sequence numbers of each sample were normalized to 2983 using the function of rarefication_even_deepness, without replacement, on the phyloseq library from R.

3.4.1.2. Chao1:

The Chao1 estimates the total number of species present in a community **(267, 345).** Before normalization, Chao1 numbers of the SaiGon river varied from 5243 to 14257, with sample SG2b3 was the highest and RF1b2 the lowest. After normalization, Chao1 numbers of the SaiGon river varied from 5243 to 9783 with sample SG6a1 was the highest and RF1b2 the lowest.

Chao1 numbers of the DongNai river before normalization varied from 9435 to 19641, with sample DN2a3 was the highest and DN3a2 the lowest. The species richness did not express any particular pattern from upstream to downstream of the river. After the normalization, Chao1 numbers of the DongNai river varied from 5948 to 11848, with sample DN2b1 is highest and that DN3a2 is lowest. The Chao1 numbers were lower in samples DN3a1, SG4b1, SG4a1, DN2a1, DN3b1, SG4b3 and DN3a2. Except for the sample DN2a1, all these samples belong to downstream locations, which are DN3 and SG4. Interestingly, before and after the normalization, the Chao1 estimators of the SaiGon river are lower than those of the DongNai river, suggesting the lower species richness of the SaiGon branch.

In summary, the behavior of the Chao1 estimators are observed to the OTU number before and after normalization in several cases, with higher Chao1 numbers in the DongNai branch compared to the SaiGon branch, and lowest numbers in samples SG8a1. In contrast with OTUs pattern, Chao1 number did not decrease at downstream of SaiGon river before and after normalization and also not decrease at downstream of DongNai river before normalization but did after normalization.

3.4.1.3. Shannon:

The Shannon index expresses the evenness diversity of a community. It means that a community numerically dominated by one or a few species is said to exhibit low evenness, while a community where abundance is equally distributed amongst species exhibits high evenness **(268).**

Before normalization, the Shannon numbers of the SaiGon river varied from 7.14 to 8.03, with sample SG1a1 is the highest and SG3a1 the lowest. After the normalization, Shannon numbers of the SaiGon river varied from 6.92 to 7.42, with sample SG1a1 is the highest and SG3b1 the lowest.

Shannon index of the DongNai river before the normalization varied from 7.25 to 8.39, with sample DN2b1 is the highest and that of DN3b1 the lowest. Species evenness decreased at the samples SG4b3, DN3a2, SG4b1, SG4a1 and DN3b1, which are downstream locations of the DongNai river, but not in samples DN3a1 and SG4b2.

After the normalization, species evenness of the DongNai river varied from 7.04 to 7.68, with sample DN2b1 is the highest and DN3b1 is lowest. Species evenness decreased clearly decreased at all the samples belonging to downstream of DongNai river which are DN3 and SG4 locations. Interestingly, before and after the normalization, Shannon indexes of the SaiGon river are lower than those of the DongNai river, suggesting lower species evenness of the SaiGon river.

Shannon indexes in canal samples, which are SG8a1 & SG9a1, were the lowest among the other samples both before and after normalization (6.80 & 6.88 and 6.74 & 6.57, respectively). The Shannon indexes of the samples from the SaiGon river did not decrease, but did decrease downstream in the DongNai river before and after normalization. The Shannon index in the DongNai river is higher than that of the SaiGon river.

3.4.2. Taxonomic assignment of bacteria at the phyla level:

3.4.2.1. Before 16S rDNA copy numbers & sequences numbers normalizing:

Sequences were classified with the Silva NGS website using the Silva 123 database release, with classification similarity set at 90% and sequence identity 1 (**Materials & Methods**). The classified sequences at the phylum level range from 90.92% to 96.65 % among 40 samples. These sequences are mainly categorized into 15 phyla, with the proportions varying among the different samples of the two river branches and the canals (**Fig. 3.25** and **Table 3.31**).

The proportion of high-quality sequences that could not be assigned to any taxa at the phylum level range from 3.35% to 9.08 %. The proportion of *Proteobacteria* was the most dominant phylum across all 42 sediment samples ranging from 10.88% to 61.63%. The second dominant phylum is *Chloroflexi*, with the proportion ranging from 8.58% to 46.43%. The phylum *Nitrospirae* was third, with the abundance ranging from 0.56% to 18.23 %.

1. Proteobacteria :

Location SG3 had highest *Proteobacteria* abundance among the samples, with the proportion of 53.16% and 61.63%. The *Proteobacteria* abundance in location SG6 ranked second among the samples. Location SG2 possessed the lowest abundance of

Proteobacteria among the samples, with the proportion ranging from 10.88% to 21.30%, subsequently. The abundance of *Proteobacteria* accumulated at the downstream of SaiGon river with the highest proportion in location SG3 and SG6.

The *Proteobacteria* abundance in the DongNai river ranged from 18.09% to 46.63%, with sample DN2b2 being the highest and DN3a2 the lowest. *Proteobacteria* abundance in the intersection location SG5 varied from 32.26% to 41.86%, which are quite similar to the *Proteobacteria* proportion in the DongNai river. The canal locations SG8a1 & SG9a1 had relatively low proportions of ***Proteobacteria*** compared with the average abundance of all the samples (28.95% & 23.26%, respectively).

2. Chlorofexi:

The second dominant phylum is *Chloroflexi* with the proportion ranging from 8.58% to 46.43%, with sample DN2b2 being the lowest and SG2b3 the highest. The *Chloroflexi* abundance dereased at the downstream of the SaiGon river with lower proportion in samples SG3 (a1, b1) and SG6 (a1, a2).

3. Nitrospirae:

Phylum *Nitrospirae* stands as the third most abundant, with the proportion ranging from 0.56% to 18.23 % with sample SG8a1 being the lowest and SG4b3 the highest. The samples SG3a1 & SG3b1 belonging to location SG3 possessed the lowest *Nitrospirae* abundance in the SaiGon river with proportion 4.43% & 6.47%, respectively. Samples SG4 (a1, b1, b2, b3), DN3b1 had the highest *Nitrospirae* proportions from 13.89% to 18.23%. The *Nitrospirae* proportions accumuated at the downstream of the DongNai river with high abundance in all samples at the SG4 location and in sample DN3b1.

The *Nitrospirae* abundance in the intersection location SG5 ranged from 5.95% to 16.47%, with sample SG5a1 with the significantly higher proportion than others. Samples from the canal locations, SG8a1 & SG9a1 had the lowest *Nitrospirae* abundance with extremely low abundance (0.56% & 0.88%). In addition, except samples of location SG3, samples from the SaiGon river had higher *Nitrospirae* proportion than that of samples from the upstream the DongNai location (RF2, DN1, DN2 and the left side of DN3).

Other phyla which were less abundant in the bacterial community of the SG-DN river system were:

- ***Acidobacteria:***

The proportion of *Acidobacteria* ranges from 1.84% to 11.61 % in all the samples. *Acidobacteria* abundance clearly decreased from the upstream to downstream of the river

with the highest proportion in all samples of RF2 location (9.49%-11.61%), decreased toward all the samples of locations DN1 & DN2 (6.49%-8.42%) and at the samples of locations DN3 & SG4 (3.31%-4.08%). The *Acidobacteria* abundance in canal locations SG9a1 (1.84%) were lowest among all the samples.

- ***Aminicenates:***

Aminicenates abundance were from 0.22% to 7.78%. *Aminicenates* expressed the abundance at the upstream of the SaiGon river with all the samples of locations RF1, SG1, SG3 ranging from 1.32%-5.01% and decreased at the samples of locations SG3 & SG6 (0.31%-0.85%).

- ***Bacteriodetes:***

Bacteriodetes abundance ranged from 0.18% to 7.66%. Samples SG6 (a1, a2) had the highest proportion of *Bacteriodetes* in the SaiGon river (3.97%, 3.62%, respectively). In DongNai river, the *Bacteriodetes* abundance were high in the samples at the upstream, including RF2 (a1, a2, b1, b2), DN2 (a1, a2, a3, b1, b2, b3) with the proportion ranging from 5.00%-7.66%. However, the *Bacteriodetes* abundance of samples of location DN1 were lower compared to samples of location RF2 and DN2 with the proportion 2.43%-3.80%. The *Bacteriodetes* abundance decreased at the locations downstream, including samples DN3 (a1, a2, b1) and SG4 (a1, b1, b2, b3) with the proportion from 0.65%-2.77%.

- ***Planctomycetes:***

Planctomycetes abundance ranged from 0.89%-5.64%. In the SaiGon river, the *Planctomycetes* abundance was higher in samples from upstream locations RF1, SG1 and SG2 with proportion from 3.55%-5.64%. The *Plantomycetes* abundance decreased in the samples of downstream locations SG3 & SG6 with the proportion 1.55%-2.59%. Oppositely, in the DongNai river, the *Plantomycetes* abundance increased at samples belonging to the downstream locations. The *Plantomycetes* proportions in samples from upstream locations RF2, DN1 and DN2 (a1, a2, a3) were from 2.28% to 3.15%. The samples from downstream locations, including DN2 (b1, b2, b3), DN3 and SG4 had higher *Plantomycetes* proportion (2.96%-4.41%). The *Plantomycetes* proportion of intersection location SG5 were similar to that of downstream of the DongNai river (2.83-4.31%). The samples of canal locations, SG8a1 had the lowest *Plantomycetes* proportions (0.89%).

- ***TA06*:**

TA06 abundance ranged from 0.07%-3.38% with sample SG2b3 is highest. In the SaiGon river, *TA06* proportion was higher at the samples belonging to upstream locations,

including RF1 (a1, b2), SG1 (a1, a2), SG2 (a1, a2, a3, b1, b2, b3); and decreased in the samples at the downstream of the river; including SG3 (a1, b1) & SG6 (a1, a2).

- ***Chlorobi:***

Chlorobi abundance ranged from 0.38% to 3.72 % with sample SG6a2 is highest. Samples of downstream locations in both river, including SG6 (a1, a2), SG4 (a1, b1, b2, b3) and DN3b1; and intersection location SG5 (a1, a2, b1, b2) had similar proportion of *Chlorobi* (from 2.74%-3.72%).

- ***Spirochaetae:***

Spirochaetae abundance ranged from 0.63% to 3.64 % among the samples. In the SaiGon river, *Spirochaetae* porportions were higher in the samples from upstream locations, inculding RF1 (a1, b2), SG1 (a1, a2), SG2 (a1, a2, a3, b1, b2, b3) (1.99%-3.51%), except lower porportions in samples RF1 (a2, b1) (1.53%-1.58%); then decreased toward samples SG3 (b1) and SG6 (a1, a2) which belonged to downstream locations (0.92%-1.53%). In the DongNai river, *Spirochaetae* porportions were highest at the samples DN3 (a1, a2) (3.46%, 3.64%).

- ***Firmicutes:***

Firmicutes abundance ranged from 0.23% to 3.64 % among the samples. *Firmicutes* abundance was highest in canal samples SG8a1 & SG9a1 with the proportion 3.26 % & 3.25 %, respectively.

- ***Elmusimicrobia*:**

Elmusimicrobia abundance ranged from 0.19% to 1.66% among the samples. In the SaiGon river, *Elmusimicrobia* abundance was higher in the samples belonging to upstream locations, including RF1 (a1, a2, b1, b2), SG1 (a1, a2), SG2 (a2, a3, b1, b2, b3) with the proportion ranging from 0.95% to 1.66%. *Elmusimicrobia* abundance was lower in the samples belonging to downstream locations, including SG3 (a1, b1) & SG6 (a1, a2) with the proportion ranging from 0.31%-0.51%.

- ***Gemmatimonadetes*:**

Gemmatimonadetes abundance was up to 1.85 % among the samples. In SaiGon river, only sample **S**G6a2 had *Gemmatimonadetes* proportion greater 1% (1.26%), compared to others. In the DongNai river, samples RF2 (a1, a2, b1, b2), DN2 (b1, b2, b3) and SG4 (a1, b1, b2) had *Gemmatimonadetes* proportion greater 1% (1.09%-1.56%). *Gemmatimonadetes* abundance ranged from 1.09%-1.85% in intersection samples SG5

(a1, a2, b1, b2). *Gemmatimonadetes* tend to accumulate at the downstream of the 2 rivers and intersection location.

3.4.2.2. After 16S copy number normalizing process:

The high-quality sequences of 42 sediment samples were adjusted for the 16S copy numbers by Tax4Fun program (described in **Materials & Methods**). The phyla proportions changed after the 16S copy numbers normalizing process (**Fig. 2.26** & **Table 3.32**).

The unclassified sequences increased significantly from 13.53% to 51.46%, compared to 3.35%-9.08% before the process. However, *Proteobacteria*, *Chloroflexi* and *Nitrospirae* remained the three most abundant phyla among all the samples. The *Proteobacteria* proportion ranged from 15.96% to 59.44% making it the most dominant phylum across all 42 sediment samples. The *Chloroflexi* proportion ranged from 5.98% to 18.32% and *Nitrospirae* ranged from 0.51% to 31.26% of all the samples. Several phyla had disappeared after normalization, including *Spirochaetae*, *Chlorobi*, *Aminicenantes*, *TA06*.

1. Proteobacteria:

The variation of *Proteobacteria* abundance before and after 16S copy number normalization ranged from -6.47% to +10.21% (calculated but not shown). The behavior pattern of *Proteobacteria* across the 42 samples after the normalization was the same with the pattern before the normalization, expect the strong variation in sample DN3a2 (10.21%).

2. Nitrospirae:

Nitrospirae abundance mostly increased after 16S copy number normalization with the proportion, variation from -0.14% to +13.84%. The behavior pattern of *Nitrospirae* across the 42 samples after the normalization was similar to the pattern before the normalization. However, the variation in the samples SG6 (a1, a2), SG5a1, SG4 (a1, a2, b1, b2) and DN3b1 increased from +7.71% to +13.84%.

3. Chloroflexi:

In contrast, *Chloroflexi* abundance decreased after the 16S copy number normalization with the proportion variation ranging from -0.86% to -34.33%. *Chloroflexi* abundance of samples from upstream the SaiGon river, including RF1a1, SG2 (a1, a2, a3, b1, b2, b3) and samples from the DongNai river, such as DN3 (a1, a2) significantly decreased from 18.41% to 34.33%. Despite the strong variation of *Chloroflexi* proportion

before and after the normalization, the behavior pattern of *Chloroflexi* abundance remained the same for all the samples, except the samples from location SG2.

3.4.2.3. After sequences numbers normalizing process:

The normalized sequences (2983 seqs/sample) were classified with the Silva NGS website using the Silva 123 database release, classification similarity at 90% and sequence identity 1. The classified sequences at the phylum level ranged from 70% to 87% among 40 samples. Those sequences were mainly categorized into 18 phyla, with the proportions varying among different samples of the two river branches and the canals (**Fig. 3.27** & **Table 3.33**).

Phyla that present in all samples belong to the *Proteobacteria* (8.4%-56.1%), *Chloroflexi* (7.4%-37.9%), *Nitrospirae* (0.4%-16.8%), *Acidobacteria* (1.0%-10.1%). Less abundant phyla that present in the samples were found belong to the *Bacteroidetes* (0.2%-6.7%), *Actinobacteria* (0.3%-4.7%), *Aminicenantes* (0.1%-6.4%), *Chlorobi* (0.3%-3.5%), *Planctomycetes* (0.5%-3.9%), *Verrucomicrobia* (1.1%-4.0%), *Spirochaetae* (0.6%-3.2%) and *Firmicutes* (0.2%-2.9%).

Nitrospirae abundance is lowest at SG8a1 & SG9a1 (0.4% and 0.5%, respectively) and highest at DN3b1, SG4a1, SG4b3, SG5a1, SG4b1 and SG4b2 (16.8 %-13.3 %, subsequently).

Acidobacteria abundance is lowest at SG2a1 & SG9a1 (1.7% and 1.0%, respectively) and highest at RF2b2, RF2b1, RF2a2 and RF2a1 (10.0%-8.4%, subsequently).

Chloroflexi abundance is lowest at SG3b1 & DN2b2 (7.4%) and highest at DN3a2, SG9a1, RF1a1 and SG8a1 (31.8 ; 32.5 ; 32.9 and 36.5 %) subsequently.

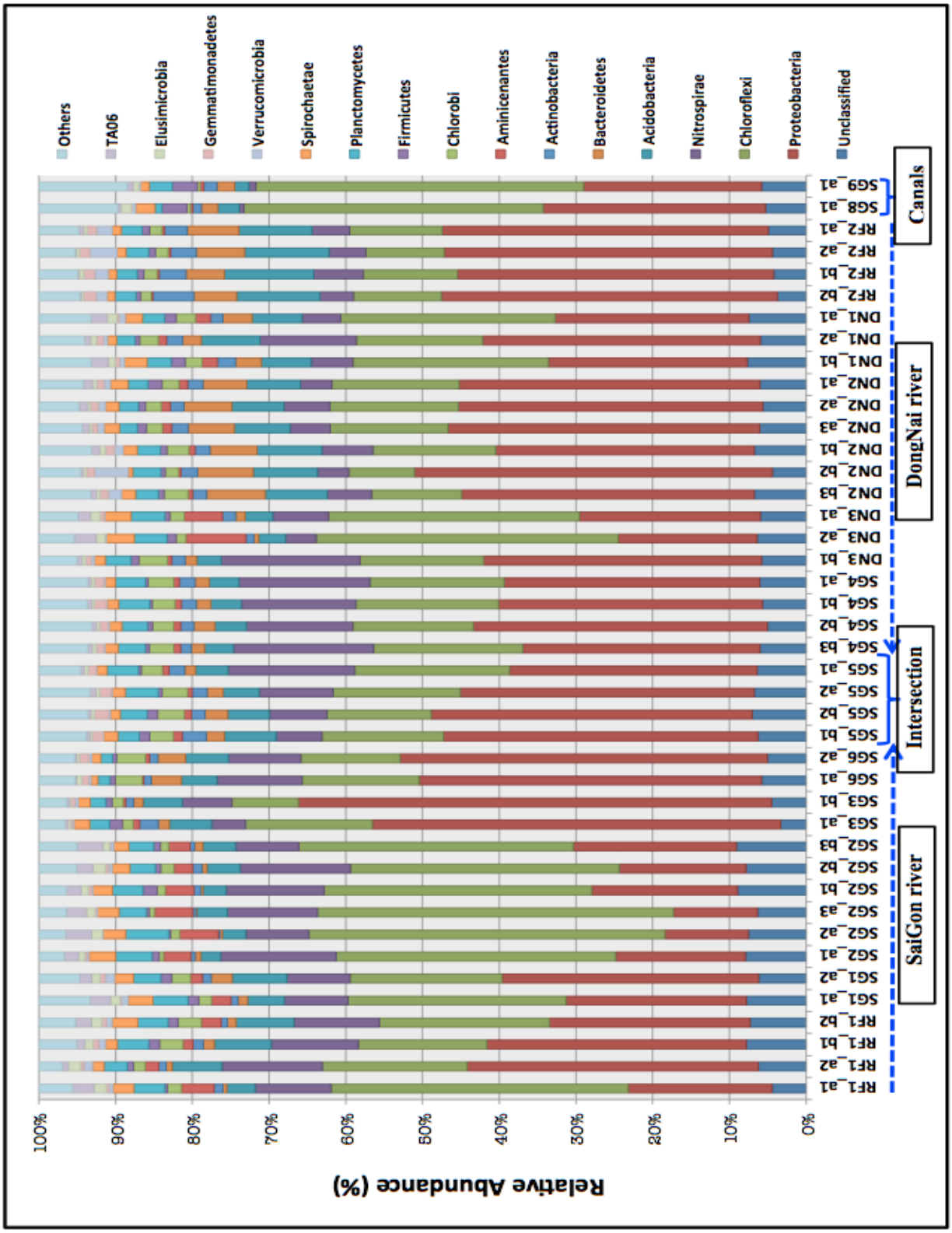


Figure 3.25. Relative abundance of the bacterial phyla populations of 42 sediment samples from the SG-DN river system before the normalization processes. On the right of the graph: SaiGon river, intersection (SG5), DongNai river and Canals (SG8 & SG9). The arrows indicate the flow of the river from upstream to downstream. These are the top 15 phyla of bacterial polulations which all have relative abundance >1%.

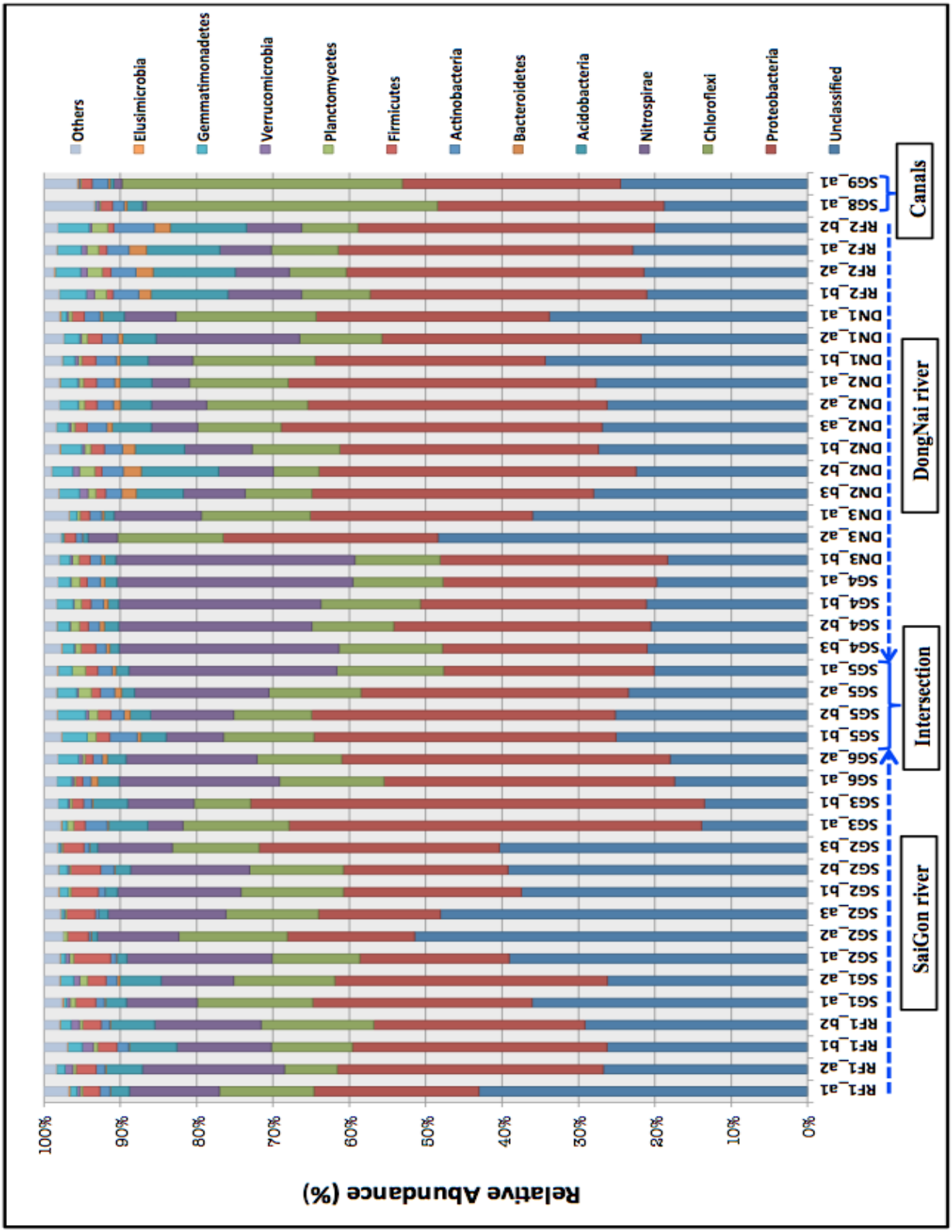


Figure 3.26. Relative abundance of the bacterial phyla populations of 42 sediment samples from the SG-DN river system after the 16S copy numbers normalization. On the right of the graph: SaiGon river, intersection (SG5), DongNai river and Canals (SG8 & SG9). The arrows indicate the flow of the river from upstream to downstream. These are the top 15 phyla of bacterial polulations which all have relative abundance >1%

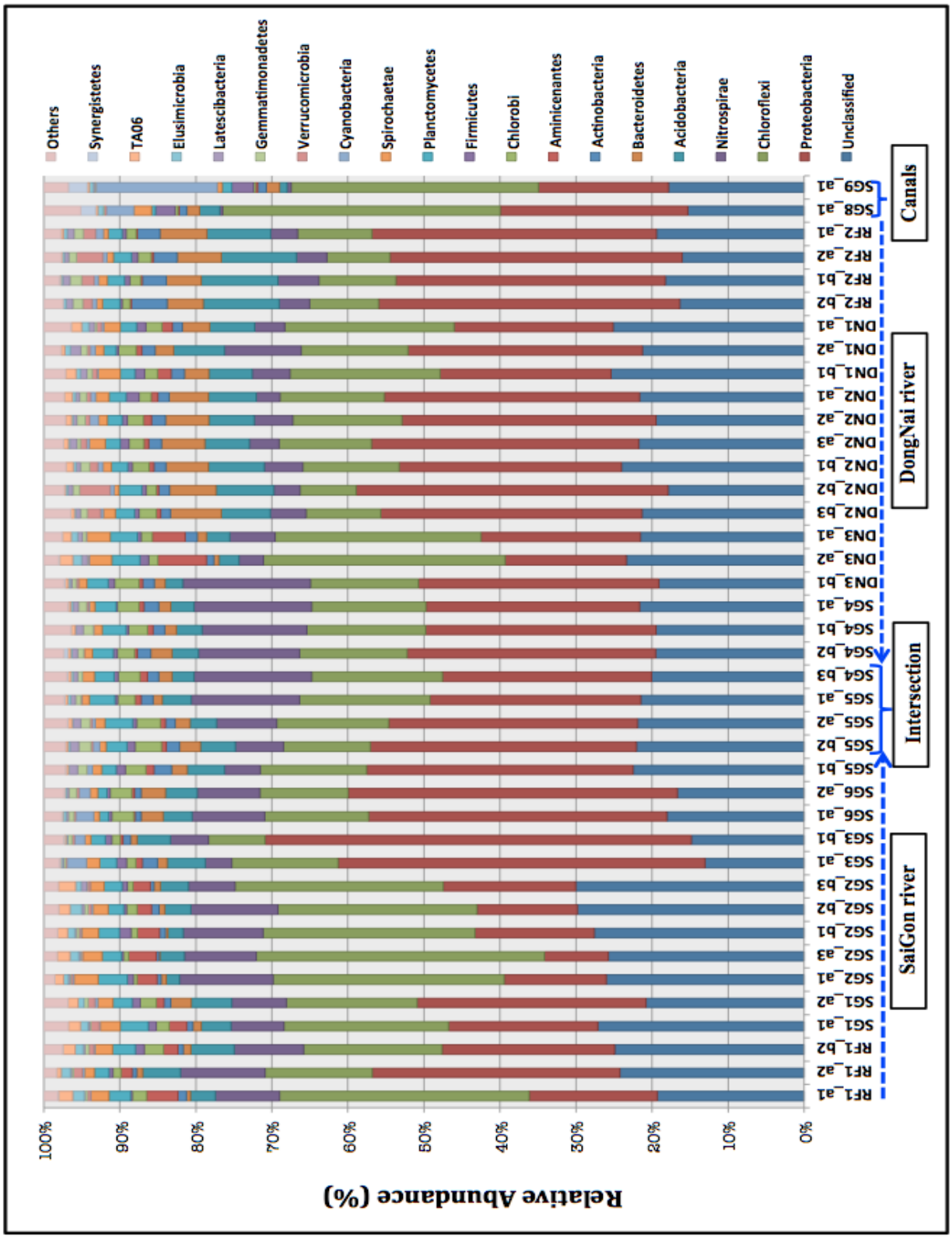


Figure 3.27. Relative abundance of the bacterial phyla populations of 40 sediment samples from the SG-DN river system after the sequence numbers normalization to 2983 seqs/ sample. Samples RF1b1 & SG2a2 were eliminated. On the right of the graph: SaiGon river, intersection (SG5), DongNai river and Canals (SG8 & SG9). The arrows indicate the flow of the river from upstream to downstream. These are the top 18 phyla of bacterial polulations which all have relative abundance >1%.

Table 3.31: Top 15 phyla that have relative abundance > 1% for total 42 sediment samples before sequence normalizing processes (including page 148).

	RF1_a1	RF1_a2	RF1_b1	RF1_b2	SG1_a1	SG1_a2	SG2_a1	SG2_a2	SG2_a3	SG2_b1	SG2_b2	SG2_b3	SG3_a1	SG3_b1	SG6_a1	SG6_a2	SG5_a1	SG5_a2	SG5_b1	SG5_b2
Unclassified	4.48	6.23	7.78	7.29	7.79	6.20	7.88	7.54	6.33	8.93	7.87	9.08	3.35	4.53	5.80	5.04	6.42	6.77	6.27	7.05
Nitrospirae	9.91	13.08	11.39	11.16	8.27	8.24	14.97	8.18	11.72	12.74	14.34	8.21	4.43	6.47	11.14	9.41	16.47	9.63	5.95	7.47
Chloroflexi	38.71	18.84	16.75	22.15	28.45	19.80	36.51	46.40	46.43	34.90	35.00	35.76	16.55	8.67	15.22	12.96	20.15	16.63	15.84	13.56
Proteobacteria	18.71	38.00	33.82	26.18	23.48	33.42	16.91	10.88	10.94	19.03	16.54	21.30	53.16	61.63	44.66	47.87	32.26	38.28	41.02	41.86
Spirochaetae	2.78	1.53	1.58	3.33	3.21	2.49	3.51	2.94	2.79	2.54	2.34	1.99	2.06	1.53	0.92	1.15	1.36	1.73	1.83	1.26
Gemmatimonadetes	0.34	0.43	0.75	0.57	0.17	0.70	0.26	0.00	0.17	0.44	0.46	0.16	0.29	0.54	0.93	1.26	1.09	1.48	1.44	1.85
Chlorobi	1.86	1.51	3.03	2.99	1.68	2.51	0.62	1.21	0.73	1.08	1.77	1.05	1.44	1.44	3.47	3.75	2.74	3.36	3.07	3.54
Acidobacteria	3.66	6.45	7.39	7.53	4.84	7.15	2.61	3.11	4.01	3.01	4.35	4.28	5.49	5.16	4.60	5.59	4.30	4.65	6.71	5.45
Planctomycetes	4.00	2.99	4.09	3.97	4.62	3.55	4.71	5.64	3.56	3.98	3.33	3.21	2.59	2.05	1.59	1.55	4.10	4.31	2.83	3.49
Bacteroidetes	0.57	0.83	1.45	1.11	1.23	2.72	0.74	0.29	0.18	0.30	0.62	1.09	1.46	1.19	3.97	3.62	1.38	2.12	2.40	2.97
Actinobacteria	1.10	0.94	1.32	0.84	0.90	1.12	0.55	0.23	0.34	0.93	1.26	0.55	2.40	1.07	0.95	1.09	2.06	1.92	3.14	1.76
Firmicutes	0.34	0.81	1.41	1.31	1.37	1.40	0.90	0.23	0.37	1.86	0.69	0.88	1.63	0.83	0.71	0.53	0.35	0.46	1.33	1.34
Aminicenantes	4.24	1.79	1.32	2.59	2.53	1.55	3.56	5.01	4.94	3.68	2.39	2.83	0.86	0.31	0.22	0.53	0.85	0.60	1.18	0.91
TA06	2.73	0.75	1.01	2.05	2.62	1.53	1.85	3.34	2.56	1.88	2.10	3.38	0.31	0.33	0.14	0.18	0.44	0.67	0.48	0.38
Verrucomicrobia	0.48	1.12	0.88	0.77	0.92	1.16	0.30	0.06	0.14	0.13	0.41	0.42	0.22	0.40	0.24	0.36	0.06	0.13	0.04	0.24
Elusimicrobia	1.66	1.63	1.06	1.38	1.24	1.11	0.79	1.44	1.18	0.95	1.59	0.84	0.34	0.19	0.69	0.40	0.50	0.51	0.31	0.47
Others	4.41	3.06	4.97	4.77	6.68	5.34	3.32	3.51	3.61	3.62	4.96	4.96	3.43	3.66	4.76	4.70	5.45	6.76	6.17	6.39

	SG4_a1	SG4_b1	SG4_b2	SG4_b3	DN3_a1	DN3_a2	DN3_b1	DN2_a1	DN2_a2	DN2_a3	DN2_b1	DN2_b2	DN2_b3	DN1_a1	DN1_a2	DN1_b1	RF2_a1	RF2_a2	RF2_ b1	RF2_b2	SG8_a1	SG9_a1
Unclassified	6.08	5.72	5.04	6.00	5.95	6.41	5.78	6.05	5.67	6.09	6.76	4.40	6.80	7.49	5.99	7.67	4.93	4.39	4.25	3.75	5.35	5.79
Nitrospirae	17.06	14.93	13.89	18.23	7.29	3.99	18.09	4.13	5.96	5.22	6.62	4.09	5.76	5.00	12.52	5.49	4.92	4.80	6.46	4.38	0.56	0.88
Chloroflexi	17.49	18.55	15.68	19.48	32.70	39.36	16.10	16.57	16.74	15.32	16.03	8.58	11.77	27.96	16.53	25.47	12.05	10.24	12.28	11.49	39.00	42.70
Proteobacteria	33.27	34.41	38.37	30.91	23.60	18.09	36.29	39.18	39.69	40.62	33.71	46.63	38.09	25.24	36.14	25.91	42.52	42.78	41.21	43.78	28.95	23.26
Spirochaetae	1.30	1.49	1.59	1.72	3.46	3.64	1.49	2.30	1.79	2.03	1.79	0.63	1.81	2.30	1.14	2.95	1.22	1.14	1.07	1.04	2.53	1.15
Gemmatimonadetes	1.19	1.56	1.15	0.94	0.53	0.12	0.77	0.96	1.04	0.72	1.16	1.09	1.09	0.30	0.88	0.65	1.27	1.33	1.46	1.67	0.06	0.09
Chlorobi	3.32	2.95	2.77	3.16	1.91	1.30	3.72	2.18	2.02	2.10	2.89	1.70	3.09	2.49	2.46	2.22	1.66	1.67	1.74	1.39	0.45	0.38
Acidobacteria	3.88	4.01	4.08	3.75	3.57	3.51	3.14	6.99	6.79	7.27	8.42	8.37	8.06	6.54	7.65	6.49	9.49	11.00	11.61	10.80	2.84	1.84
Planctomycetes	3.95	4.04	3.35	3.56	4.41	4.24	3.31	2.59	2.57	2.28	3.02	3.66	2.96	2.85	2.43	3.15	2.75	2.99	2.68	2.71	0.89	3.01
Bacteroidetes	1.88	1.86	2.77	1.75	1.27	0.65	1.54	5.73	6.26	6.04	6.14	7.27	7.66	3.80	2.43	3.34	6.78	6.26	5.00	5.59	2.12	2.35
Actinobacteria	2.00	1.93	1.67	1.41	1.74	1.00	1.75	1.98	1.79	2.15	1.99	2.13	1.84	1.61	2.09	2.37	2.93	3.34	3.52	5.33	1.14	1.68
Firmicutes	0.32	0.34	0.62	0.51	0.61	1.23	1.01	1.86	0.90	1.22	0.85	0.60	0.76	1.51	0.59	1.79	0.93	0.87	0.88	0.60	3.26	3.25
Aminicenantes	0.78	0.89	0.93	0.91	4.93	7.78	0.58	1.14	1.17	1.19	0.77	0.33	0.60	1.97	1.06	2.00	0.32	0.32	0.26	0.25	0.31	0.48
TA06	0.34	0.57	0.49	0.58	1.64	2.82	0.60	1.18	1.12	0.94	1.06	0.31	0.68	2.07	0.77	2.20	0.57	0.18	0.19	0.07	0.36	0.67
Verrucomicrobia	0.18	0.10	0.17	0.17	0.09	0.04	0.34	0.84	0.89	0.82	1.17	4.36	1.75	0.78	0.54	0.62	1.96	3.59	1.70	1.39	0.47	0.10
Elusimicrobia	0.51	0.32	0.45	0.53	1.21	1.15	0.50	0.54	0.34	0.37	0.76	0.39	0.49	1.29	0.81	0.83	0.49	0.41	0.64	0.35	1.45	0.82
Others	6.43	6.34	6.97	6.38	5.11	4.66	4.99	5.81	5.25	5.64	6.87	5.46	6.78	6.79	5.96	6.85	5.22	4.69	5.05	5.40	10.27	11.56

Table 3.32: Top 15 phyla that have relative abundance > 1% for total 42 sediment samples after 16S copy numbers normalization (including page 150).

	RF1_a1	RF1_a2	RF1_b1	RF1_b2	SG1_a1	SG1_a2	SG2_a1	SG2_a2	SG2_a3	SG2_b1	SG2_b2	SG2_b3	SG3_a1	SG3_b1	SG6_a1	SG6_a2	SG5_a1	SG5_a2	SG5_b1	SG5_b2
Unclassified	43.07	26.76	26.26	29.17	36.11	26.21	39.09	51.46	48.11	37.48	39.26	40.35	13.91	13.53	17.42	18.03	20.13	23.51	25.10	25.20
Nitrospirae	11.81	18.61	12.42	13.96	9.28	9.53	19.07	10.67	15.45	16.17	15.60	9.72	4.63	8.68	21.00	17.12	27.21	17.61	7.49	10.90
Chloroflexi	12.35	6.92	10.59	14.74	15.12	13.23	11.43	14.23	12.10	13.42	12.24	11.36	13.88	7.41	13.66	11.11	13.99	12.03	11.89	10.20
Proteobacteria	21.61	34.86	33.37	27.68	28.71	35.71	19.59	16.68	15.96	23.32	21.56	31.52	54.02	59.44	38.09	42.99	27.56	35.00	39.54	39.78
Spirochaetae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gemmatimonadetes	0.93	1.10	1.90	1.47	0.46	1.76	0.73	0.00	0.46	1.15	1.15	0.32	0.72	1.32	1.95	2.75	1.82	2.53	3.32	3.69
Chlorobi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Acidobacteria	2.44	4.74	6.20	5.76	2.70	5.37	1.24	0.74	1.18	1.60	2.06	1.09	5.10	4.57	2.78	2.50	1.70	1.70	3.31	2.65
Planctomycetes	0.33	0.43	0.59	0.46	0.78	1.04	0.58	0.58	0.27	0.41	0.22	0.25	0.83	0.33	0.30	0.34	1.80	1.69	1.14	1.25
Bacteroidetes	0.09	0.19	0.08	0.09	0.18	0.39	0.15	0.08	0.01	0.03	0.07	0.08	0.18	0.11	0.94	0.61	0.52	0.90	0.46	0.82
Actinobacteria	1.35	1.12	1.50	1.14	1.12	1.44	0.74	0.33	0.47	0.91	1.80	0.68	2.90	1.10	1.06	1.24	1.89	1.91	3.61	1.72
Firmicutes	2.25	2.59	2.55	2.41	2.63	2.42	4.79	2.75	3.76	3.51	3.92	2.70	1.51	1.49	0.91	1.04	1.58	1.18	1.79	1.71
Aminicenantes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TA06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Verrucomicrobia	0.40	1.04	1.45	1.04	0.39	0.77	0.48	0.00	0.04	0.00	0.21	0.14	0.00	0.19	0.22	0.48	0.00	0.25	0.00	0.35
Elusimicrobia	0.20	0.08	0.06	0.09	0.18	0.11	0.02	0.00	0.10	0.17	0.05	0.05	0.18	0.00	0.00	0.03	0.16	0.12	0.08	0.09
Others	3.17	1.57	3.03	2.00	2.33	2.01	2.07	2.48	2.10	1.84	1.85	1.72	2.14	1.83	1.68	1.76	1.65	1.57	2.26	1.65

	SG4_a1	SG4_b1	SG4_b2	SG4_b3	DN3_a1	DN3_a2	DN3_b1	DN2_a1	DN2_a2	DN2_a3	DN2_b1	DN2_b2	DN2_b3	DN1_a1	DN1_a2	DN1_b1	RF2_a1	RF2_a2	RF2_b1	RF2_b2	SG8_a1	SG9_a1
Unclassified	19.77	21.11	20.53	21.00	36.04	48.43	18.35	27.74	26.30	26.95	27.40	22.45	28.05	33.85	21.81	34.38	22.91	21.45	21.05	20.05	18.84	24.53
Nitrospirae	30.90	26.48	25.31	28.76	11.43	3.85	31.26	4.93	7.22	6.05	8.86	7.19	8.07	6.77	18.83	5.81	6.72	7.04	9.69	7.26	0.51	1.08
Chloroflexi	11.77	13.04	10.75	13.58	14.29	13.82	11.16	12.94	13.24	10.96	11.43	5.98	8.77	18.32	10.76	16.03	8.78	7.46	8.94	7.43	38.14	36.73
Proteobacteria	28.01	29.62	33.65	26.82	29.09	28.13	29.81	40.30	39.17	41.99	33.91	41.56	36.88	30.56	33.97	30.14	38.56	38.99	36.25	38.79	29.63	28.55
Spirochaetae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gemmatimonadetes	1.67	2.21	1.63	1.57	1.02	0.34	1.35	2.33	2.43	1.61	2.80	2.74	2.68	0.77	2.03	1.61	3.21	3.30	3.54	4.13	0.14	0.13
Chlorobi	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
Acidobacteria	1.63	1.40	1.84	1.25	1.35	0.65	1.49	4.14	4.11	5.13	6.50	10.09	6.18	2.82	4.34	3.73	9.62	10.81	10.04	9.96	2.03	0.46
Planctomycetes	1.11	0.96	1.17	0.81	0.46	0.11	0.89	0.61	0.76	0.58	0.80	2.01	1.05	0.47	0.83	0.50	1.55	1.97	1.56	2.11	0.17	0.15
Bacteroidetes	0.55	0.54	0.66	0.48	0.28	0.12	0.45	0.70	0.87	0.74	1.66	2.37	1.89	0.41	0.63	0.48	2.30	2.23	1.62	2.07	0.41	0.25
Actinobacteria	1.74	1.67	1.44	1.32	1.55	0.85	1.46	2.40	2.19	2.57	2.34	2.77	2.09	2.02	2.08	2.67	2.89	3.27	3.36	5.33	1.46	2.13
Firmicutes	0.98	1.27	1.20	1.95	1.23	1.50	1.44	1.66	1.61	1.57	1.76	0.95	1.30	1.62	1.87	1.81	1.04	1.14	0.88	0.80	1.58	1.44
Aminicenantes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TA06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Verrucomicrobia	0.10	0.07	0.13	0.11	0.00	0.00	0.31	0.16	0.08	0.23	0.40	0.89	1.08	0.20	0.24	0.42	0.73	0.87	1.06	0.29	0.43	0.09
Elusimicrobia	0.02	0.02	0.03	0.09	0.04	0.04	0.08	0.16	0.04	0.06	0.15	0.02	0.07	0.19	0.08	0.11	0.06	0.20	0.06	0.03	0.11	0.15
Others	1.76	1.61	1.64	2.25	3.23	2.15	1.94	1.93	1.98	1.56	1.97	0.99	1.89	2.01	2.52	2.32	1.61	1.28	1.93	1.76	6.55	4.30

Table 3.33: Top 18 phyla that have relative abundance > 1% for total 40 sediment samples after sequence numbers normalization (including page 152).

	RF1_a1	RF1_a2	RF1_b2	SG1_a1	SG1_a2	SG2_a1	SG2_a3	SG2_b1	SG2_b2	SG2_b3	SG3_a1	SG3_b1	SG6_a1	SG6_a2	SG5_a1	SG5_a2	SG5_b1	SG5_b2
Unclassified	19.31	24.20	24.91	27.15	20.80	26.05	25.75	27.59	29.80	30.04	13.04	14.82	18.04	16.66	21.49	21.92	22.49	22.06
Proteobacteria	16.86	32.58	22.73	19.61	30.10	13.38	8.41	15.72	13.21	17.40	48.27	56.08	39.22	43.31	27.69	32.72	35.07	35.03
Chloroflexi	32.85	14.11	18.17	21.66	17.20	30.41	37.88	27.86	26.18	27.42	13.98	7.44	13.68	11.60	17.16	14.75	13.98	11.36
Nitrospirae	8.45	11.16	9.15	6.97	7.20	12.34	9.49	10.49	11.46	6.13	3.45	5.03	9.55	8.25	14.28	7.91	4.69	6.34
Acidobacteria	3.25	4.96	5.73	3.92	5.30	1.71	3.18	2.01	3.45	3.69	5.03	4.36	3.82	4.16	3.79	3.49	4.89	4.63
Bacteroidetes	0.50	0.77	1.01	1.07	2.70	0.70	0.17	0.30	0.70	0.87	1.27	0.84	2.92	3.25	1.21	1.91	2.08	2.75
Actinobacteria	1.14	0.57	0.64	0.80	0.90	0.54	0.30	0.80	1.04	0.47	2.01	1.04	0.80	0.84	1.61	1.31	2.35	1.71
Aminicenantes	4.12	1.51	1.91	2.31	1.00	2.61	3.69	2.88	1.88	2.25	0.84	0.34	0.13	0.37	0.77	0.67	1.01	0.67
Chlorobi	1.88	1.04	2.55	1.71	2.10	0.50	0.67	0.87	1.37	0.74	1.21	1.14	2.88	2.85	2.28	3.12	2.72	3.45
Firmicutes	0.23	0.54	1.14	1.04	1.10	0.77	0.23	1.51	0.40	0.70	1.31	0.80	0.44	0.40	0.37	0.50	1.21	1.04
Planctomycetes	2.82	1.91	3.02	3.69	2.50	3.86	2.55	2.78	2.04	2.38	2.25	1.91	1.21	1.21	3.39	3.65	1.91	2.75
Spirochaetae	2.38	1.27	2.35	2.68	2.00	3.18	2.55	2.21	2.01	1.81	1.74	0.80	0.77	1.04	0.97	1.27	1.21	0.87
Cyanobacteria	0.07	0.30	0.17	0.10	0.40	0.17	0.00	0.17	0.20	0.03	2.58	1.37	2.38	1.48	0.10	0.44	0.74	0.87
Verrucomicrobia	0.44	1.14	0.64	1.11	0.90	0.23	0.17	0.20	0.40	0.34	0.17	0.30	0.20	0.30	0.03	0.23	0.03	0.27
Gemmatimonadetes	0.34	0.27	0.50	0.13	0.50	0.10	0.13	0.34	0.50	0.07	0.23	0.60	0.70	1.01	0.67	1.24	1.17	1.58
Latescibacteria	0.03	0.17	0.23	0.23	0.00	0.23	0.20	0.20	0.37	0.74	0.10	0.03	0.34	0.27	0.64	1.07	1.11	1.21
Elusimicrobia	1.51	1.21	1.04	0.97	0.80	0.64	1.21	0.94	1.51	0.77	0.23	0.17	0.40	0.20	0.44	0.10	0.10	0.23
TA06	1.91	0.57	1.61	1.58	1.30	1.17	1.64	1.31	1.48	2.25	0.17	0.23	0.10	0.10	0.37	0.54	0.37	0.30
Synergistetes	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.00
Others	1.88	1.71	2.48	3.25	3.20	1.41	1.78	1.81	1.98	1.91	2.04	2.61	2.38	2.68	2.72	3.12	2.85	2.88

	SG4_a1	SG4_b1	SG4_b2	SG4_b3	DN3_a1	DN3_a2	DN3_b1	DN2_a1	DN2_a2	DN2_a3	DN2_b1	DN2_b2	DN2_b3	DN1_a1	DN1_a2	DN1_b1	RF2_a1	RF2_a2	RF2_b1	RF2_b2	SG8_a1	SG9_a1
Unclassified	21.62	19.48	19.51	20.08	21.56	23.40	19.11	21.62	19.48	21.72	24.00	17.87	21.32	25.11	21.25	25.38	19.44	16.02	18.27	16.36	15.29	17.83
Proteobacteria	28.13	30.31	32.69	27.49	20.95	15.92	31.61	33.59	33.42	35.20	29.27	41.07	34.36	20.92	30.84	22.49	37.41	38.45	35.43	39.62	24.67	17.13
Chloroflexi	15.02	15.62	14.15	17.20	27.05	31.81	14.21	13.71	14.38	12.14	12.67	7.38	9.82	22.29	14.08	19.75	9.76	8.31	10.12	9.05	36.54	32.48
Nitrospirae	15.45	13.81	13.31	15.49	6.00	3.22	16.80	3.15	5.06	3.92	5.06	3.45	4.69	3.96	10.06	4.96	3.55	3.99	5.36	4.02	0.37	0.54
Acidobacteria	3.08	3.35	3.52	2.92	3.05	2.61	2.31	6.27	5.93	5.83	7.34	7.58	6.47	5.90	6.70	5.70	8.41	9.86	10.09	9.96	2.61	1.04
Bacteroidetes	1.61	1.51	2.78	1.71	1.21	0.67	1.44	5.20	5.73	5.70	5.60	6.13	6.70	3.62	2.45	3.25	6.17	5.83	4.63	4.76	1.71	1.71
Actinobacteria	1.91	1.58	1.71	1.48	1.61	0.97	1.51	1.48	1.81	1.68	1.58	1.44	1.27	1.27	1.74	1.74	2.92	3.12	3.15	4.69	1.01	1.14
Aminicenantes	0.67	0.70	0.40	1.01	4.32	6.37	0.50	0.91	1.07	0.67	0.64	0.30	0.57	1.37	0.70	1.78	0.20	0.34	0.20	0.23	0.17	0.27
Chlorobi	2.82	2.48	2.25	2.85	1.48	1.24	3.18	1.58	2.11	1.94	2.21	1.31	2.25	2.11	2.35	1.71	1.31	1.71	1.44	0.97	0.37	0.27
Firmicutes	0.20	0.34	0.54	0.54	0.47	1.17	0.84	1.68	0.67	1.14	0.60	0.57	0.60	1.24	0.30	1.24	0.50	0.84	0.67	0.34	2.58	2.88
Planctomycetes	2.78	3.15	2.75	2.58	3.59	3.65	2.82	2.28	1.98	1.98	2.15	3.05	2.55	2.18	1.64	1.94	1.84	2.41	2.28	2.28	0.50	1.27
Spirochaetae	0.70	1.07	1.04	1.58	3.08	2.98	1.11	1.74	1.14	2.11	1.14	0.60	1.54	2.15	1.11	2.92	0.64	0.80	1.14	0.77	2.35	0.60
Cyanobacteria	0.20	0.07	0.00	0.03	0.07	0.03	0.03	0.67	1.17	0.37	0.67	0.54	0.40	0.27	0.67	0.17	1.01	0.57	0.60	0.57	3.62	15.99
Verrucomicrobia	0.20	0.03	0.13	0.00	0.03	0.10	0.27	0.54	0.60	0.80	1.04	3.99	1.71	0.70	0.40	0.64	1.68	3.45	1.64	1.24	0.34	0.03
Gemmatimonadetes	1.07	1.34	0.77	0.60	0.44	0.07	0.54	0.91	1.07	0.47	1.24	0.94	0.91	0.27	0.84	0.67	1.21	1.01	1.51	1.31	0.07	0.10
Latescibacteria	0.70	1.01	0.74	0.60	0.64	0.80	0.30	0.50	0.50	0.94	0.54	0.60	0.67	0.70	1.37	0.84	0.91	0.60	0.94	0.91	0.00	0.20
Elusimicrobia	0.27	0.03	0.23	0.23	0.87	1.14	0.20	0.44	0.20	0.17	0.37	0.30	0.13	0.97	0.74	0.60	0.47	0.27	0.23	0.37	0.70	0.54
TA06	0.30	0.54	0.37	0.47	1.14	1.78	0.44	1.04	0.80	0.64	0.97	0.20	0.50	1.37	0.47	1.34	0.30	0.10	0.17	0.00	0.27	0.23
Synergistetes	0.07	0.03	0.47	0.07	0.03	0.10	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.07	0.00	1.98	2.48
Others	3.18	3.55	2.65	3.08	2.41	1.94	2.72	2.72	2.85	2.58	2.92	2.68	3.52	3.52	2.28	2.88	2.28	2.31	2.04	2.55	4.86	3.25

3.4.3. Taxonomy assignment of bacteria at the genus level:

3.4.3.1. Before sequencing normalization process:

The 42 samples high-quality sequences were assigned to the genus level. A total of 16 top abundant genera were chosen to analyze. The proportion of the sequences that could not be assigned to any taxa at the genus level ranged from 4.91% to 20.01%. Interestingly, the proportion of sequences that were assigned to uncultured bacteria were high across the 42 samples, with the proportion from ranging 34.62% to 58.10% (**Table 3.34**)

3.4.3.2. After sequence number normalization process:

A total of 40 samples remained after normalizing the numbers of sequences to 2983 seqs for each sample with samples RF1b1 and SG2a1 were removed due to their sequence numbers < 2983 (**Table 3.35**). The uncultured genera included *Uncultured Anaerolineaceae, Uncultured Nitrosomonadaceae* and *Uncultured Nitrospiraceae*.

The *Uncultured Anaerolineaceae* appeared with the highest abundance in all the samples ranging from 4.19% to 14.95%. Samples of canal locations, including SG8a1 & SG9a1 had highest abundance of *Uncultured Anaerolineaceae* (14.95% & 13.85%, respectively). *Uncultured Nitrospiraceae* abundance ranged from 0.20% to 10.49%. In the SaiGon river, *Uncultured Nitrospiraceae* proportions were higher in the samples from upstream locations RF1 (a1, a2, b2), SG1 (a1, a2) with the proportion ranging from 3.99% to 5.26% and SG2 (a1, a3, b1, b2, b3) with ranging from 4.89% to 10.49%. Samples from upstream DongNai river, such as RF2 (a1, a2, b1, b2), and canal locations SG8a1 & SG9a1 had lowest *Uncultured Nitrospiraceae* abundance among the samples with the proportion < 1%. Other samples had lowest *Uncultured Nitrospiraceae* proportion 1.27% to 4.22% (**Fig. 3.28**).

Other genera such as *Nitrospira* had abundance ranging from 0.03% to 12.67%, with samples SG4a1 and DN3b1 are the highest (12.67% and 12.20%, respectively). *Anaeromyxobacter* abundance was highest in the samples SG3 (a1, b1) with the proportions 4.09% & 6.13%, respectively. The rest of the samples had *Anaeromyxobacter* proportions less than 2.08%. *Leptolinea* abundance was highest in the samples SG8a1 & SG9a1 with the significant proportion 11.50% & 7.04%, respectively. The rest of the samples had *Leptolinea* proportion less than 0.60 %, except for sample SG3a1 (1.21%).

Other uncultured genera members such as *Uncultured Alcaligenaceae* which had the highest abundance in the samples SG3 (a1, b1) (6.24% & 8.31%, respectively). The rest of the samples had *Uncultured Alcaligenaceae* proportion < 2.15%. Samples of

downstream SaiGon river, including SG3 (a1, b1) & SG6 (a1, b1) had *Uncultured Rhodocyclaceae* proportion from 1.91% to 3.92%, while others were < 1%. Samples SG8a1 also had significant *Uncultured Rhodocyclaceae* proportion (2.04%) compared to other samples.

Acidiferrobacter abundance had significant proportions in samples of the intersection location, SG5 (a1, a2, b1, b2), samples downstream of DongNai river SG4 (a1, b1, b2, b3) a,d DN3b1 (1.41%-2.72%). *Spirochaeta* appears among the samples in low abundance ranging 0.20%-3.02% and varied among the samples. In the SaiGon river, samples from upstream locations, including RF1 (a1, b2), SG1 (a1, a2), SG2 (a1, a3, b1, b2, b3), had higher *Spirochaeta* proportions than those of samples from downstream (1.61%-3.02%). *Dechloromonas* abundance was highest in the samples of the downstream SaiGon river, including samples SG3 (a1, b1) & SG6 (a1, a2) and canal sample SG8a1 with proportion from 3.55% to 4.43%.

Uncultured Hydrogenophilaceae abundance was from 1.07% to 1.24% in samples of downstream SaiGon river, including SG6 (a1, a2), samples of downstream DongNai river such as DN3b1 & SG4 (a1, b1, b2, b3), samples from intersection location such as SG5 (a1, a2, b1, b2) and canal sample SG9a1. The *Uncultured Hydrogenophilaceae* abundance showed that there is an accumulation of *Uncultured Hydrogenophilaceae* in downstream of the SaiGon and the DongNai river as well as the intersection location. *Acidobacteriaceae* (*Subgroup 1*) abundance was highest in the samples belonging to the upstream DongNai river, samples RF2 (a1, a2, b1 b2) with proportion ranging from 1.78% to 2.35%. *Acidobacteriaceae (Subgroup 1)* abundance in other samples was < 0.91%, except samples RF1a2 (1.07%) & RF1b2 (1.11%).

Longilinea abundance were < 0.37 % in all the samples except canal samples SG8a1 & SG9a1 with a relatively high proportion of 4.02% & 2.25%, respectively.
Uncultured Comamonadaceae abundance ranged from 0.03% to 2.68% in all the samples. In the SaiGon river, *Uncultured Comamonadaceae* proportions were < 1% of the samples, except the samples SG3b1 (1.01%), SG6a2 (1.61%).

Samples of canal locations, including SG8a1 & SG9a1 had the highest abundance of *Uncultured Anaerolineaceae*, *Leptolinea* and *Longilinea* among the samples. Samples of location SG3, including SG3 (a1, b1) had the highest abundance of genera *Anaeromyxobacter* and *Uncultured Alcaligenaceae* among the samples,

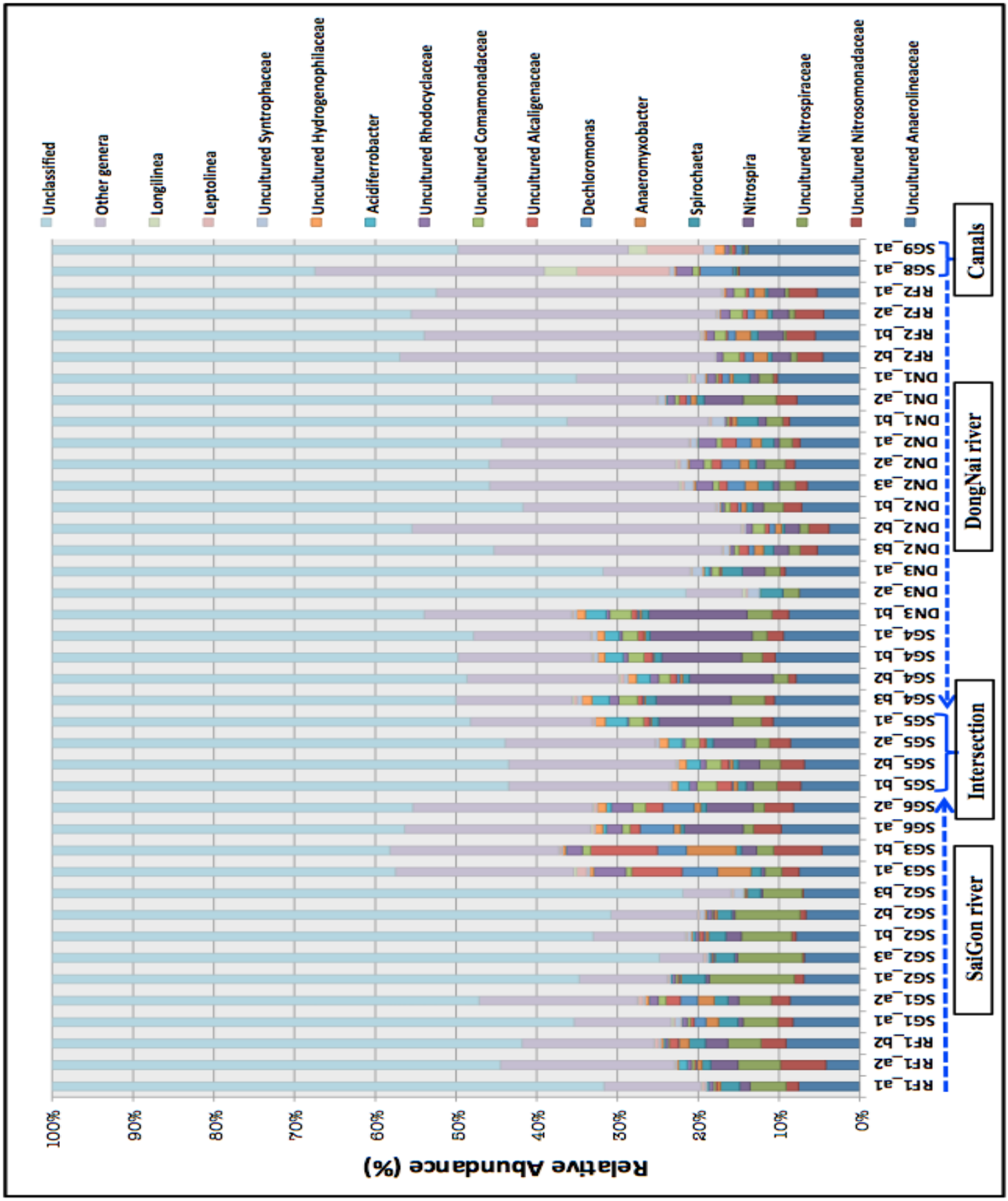


Figure 3.28. Relative abundance of the bacterial genera populations of 40 sediment samples from the SaiGon-DongNai river system. On the right of the graph: SaiGon river, Intersection (SG5), DongNai river and Canals (SG8 & SG9). The arrows indicate the flow of the river from upstream to downstream. These are the top 15 phyla of bacterial polulations which all have relative abundance >1%.

Table 3.34: Top 15 genera that have relative abundance >1% for total 42 sediment samples before sequence normalizing processes (including page 157).

	RF1_a1	RF1_a2	RF1_b1	RF1_b2	SG1_a2	SG1_a3	SG2_a1	SG2_a2	SG2_a3	SG2_b1	SG2_b2	SG2_b3	SG3_a1	SG3_b1	SG6_a1	SG6_a2	SG5_a1	SG5_a2	SG5_b1	SG5_b2
Uncultured Anaerolineaceae	8.04	4.96	7.56	10.29	10.17	9.10	7.99	8.81	7.78	9.61	8.17	7.52	8.84	5.35	10.73	8.66	12.02	9.60	8.80	7.85
Uncultured Nitrosomonadaceae	1.33	5.66	4.84	3.29	1.60	2.85	1.27	0.17	0.41	0.74	0.78	0.14	2.54	6.18	3.55	4.04	1.53	2.58	2.92	3.74
Uncultured Nitrospiraceae	5.46	6.49	5.98	5.28	5.47	5.08	12.75	6.62	9.85	8.30	10.30	6.74	2.95	3.28	1.86	2.29	4.60	2.52	3.64	3.11
Nitrospira	3.44	6.52	3.02	5.05	1.19	1.89	1.11	0.66	0.98	3.73	1.16	0.46	0.56	2.73	11.53	9.02	15.82	10.36	1.67	4.69
Spirochaeta	6.76	1.81	1.84	4.63	5.65	3.22	6.65	7.22	7.30	4.81	4.88	4.82	2.17	1.63	0.72	1.07	1.88	2.27	2.50	1.56
Anaeromyxobacter	1.07	1.02	4.72	2.38	3.59	3.64	0.40	0.49	0.33	1.04	0.65	0.86	6.80	9.00	1.31	1.24	0.24	0.29	1.06	0.78
Dechloromonas	0.00	0.15	0.15	0.30	2.89	3.90	0.25	0.00	0.22	0.26	0.23	0.25	6.41	5.35	5.58	6.15	0.14	0.03	0.19	0.27
Uncultured Alcaligenaceae	0.36	0.22	1.41	1.01	0.67	1.70	0.19	0.00	0.14	0.51	0.22	0.09	6.42	8.84	1.54	2.11	0.91	0.64	1.85	0.87
Uncultured Comamonadaceae	0.25	0.39	0.66	0.34	0.38	0.81	0.25	0.00	0.03	0.15	0.09	0.09	0.74	1.04	1.32	1.60	1.80	1.83	2.27	2.04
Uncultured Rhodocyclaceae	0.64	0.62	0.66	0.37	0.62	1.02	0.53	0.06	0.13	0.40	0.67	0.01	4.10	2.07	2.15	2.57	0.21	0.43	1.07	0.75
Acidiferrobacter	0.67	2.17	0.52	0.49	0.10	0.16	0.40	0.66	0.73	0.78	0.31	0.04	0.07	0.22	0.97	1.00	4.66	3.00	2.50	3.74
Uncultured Hydrogenophilaceae	0.02	0.29	0.13	0.20	0.04	0.26	0.05	0.00	0.01	0.06	0.24	0.03	0.55	0.36	0.90	1.24	1.36	1.27	0.85	1.01
Uncultured Syntrophaceae	0.39	0.42	0.26	0.50	0.74	0.34	0.32	0.63	0.56	0.80	0.92	1.56	0.31	0.23	0.19	0.31	0.27	0.42	0.09	0.30
Leptolinea	0.75	0.17	0.52	0.97	0.54	0.80	0.25	0.82	0.58	0.30	0.42	0.50	1.93	0.39	0.60	0.60	0.34	0.08	0.27	0.09
Longilinea	0.12	0.02	0.26	0.07	0.23	0.15	0.05	0.00	0.05	0.19	0.04	0.10	0.41	0.10	0.19	0.16	0.00	0.00	0.02	0.03
Other genera	59.29	58.25	54.41	51.62	50.44	54.51	51.46	52.34	53.32	50.01	53.47	52.08	50.29	46.71	48.51	50.37	43.76	52.84	59.38	57.09
Unclassified	11.42	10.85	13.05	13.22	15.67	10.54	16.07	21.51	17.59	18.30	17.47	24.71	4.91	6.51	8.33	7.58	10.48	11.84	10.93	12.08

	SG4_a1	SG4_b1	SG4_b2	SG4_b3	DN3_a1	DN3_a2	DN3_b1	DN2_a1	DN2_a2	DN2_a3	DN2_b1	DN2_b2	DN2_b3	DN1_a1	DN1_a2	DN1_b1	RF2_a1	RF2_a2	RF2_b1	RF2_b2	SG8_a1	SG9_a1
Uncultured Anaerolineaceae	10.42	11.19	8.66	11.36	10.43	9.00	9.55	8.78	8.92	7.55	8.25	3.85	6.15	12.33	8.49	10.83	6.20	5.30	6.65	5.42	16.06	17.66
Uncultured Nitrosomonadaceae	2.27	1.63	1.27	1.17	0.67	0.04	2.28	1.28	1.42	1.47	2.39	2.36	2.01	0.60	2.68	0.70	3.83	4.09	3.81	3.27	0.22	0.10
Uncultured Nitrospiraceae	2.35	3.14	2.43	6.00	2.69	2.44	3.53	2.20	3.18	2.75	3.53	1.37	2.08	2.41	4.86	2.46	1.22	1.14	0.73	0.93	0.31	0.46
Nitrospira	21.47	16.54	16.68	15.88	6.86	0.24	19.13	1.22	1.91	1.59	3.02	3.14	3.54	2.53	9.58	1.88	3.05	3.30	5.47	3.65	0.03	0.25
Spirochaeta	1.66	1.97	2.21	2.39	6.86	10.68	1.80	3.39	2.44	2.96	2.33	0.61	2.44	3.94	1.63	5.45	1.12	1.24	1.25	1.13	0.52	0.59
Anaeromyxobacter	0.33	0.33	0.61	0.27	0.30	0.24	0.51	2.37	2.36	3.20	1.27	1.95	2.04	1.26	1.66	1.19	1.90	2.61	2.99	2.65	0.28	0.10
Dechloromonas	0.12	0.11	0.65	0.29	0.20	0.00	0.40	3.70	3.86	4.07	0.67	1.03	1.32	1.74	0.94	0.50	1.29	1.31	1.53	1.58	5.35	1.33
Uncultured Alcaligenaceae	0.52	1.09	0.75	0.69	0.24	0.00	0.62	1.64	1.43	1.40	0.77	0.66	0.98	0.49	1.02	0.26	0.43	0.55	0.45	0.63	0.11	0.55
Uncultured Comamonadaceae	1.93	2.01	1.51	2.20	0.81	0.04	2.42	0.78	0.93	0.84	0.78	1.49	0.78	0.30	0.63	0.31	1.39	1.90	1.84	2.22	0.95	0.29
Uncultured Rhodocyclaceae	0.34	0.69	1.22	1.06	0.20	0.03	0.53	2.30	2.44	2.03	0.41	0.70	0.55	0.89	0.79	0.29	0.96	1.10	1.15	0.97	2.51	0.79
Acidiferrobacter	3.95	3.83	3.07	3.62	1.48	0.12	3.88	0.28	0.18	0.21	0.26	0.00	0.12	0.10	0.32	0.05	0.00	0.03	0.02	0.00	0.00	0.02
Uncultured Hydrogenophilaceae	0.99	1.02	1.03	1.19	0.24	0.00	1.27	0.10	0.16	0.18	0.13	0.16	0.07	0.16	0.10	0.08	0.21	0.16	0.15	0.02	0.28	1.44
Uncultured Syntrophaceae	0.72	0.70	0.76	0.61	1.37	1.76	0.43	0.92	0.81	1.32	0.43	0.25	0.79	1.47	0.81	1.95	0.08	0.30	0.16	0.09	0.50	1.79
Leptolinea	0.15	0.16	0.89	0.80	0.20	0.52	0.48	0.56	0.47	0.43	0.32	0.38	0.30	1.59	0.21	0.48	0.29	0.14	0.16	0.14	14.67	12.37
Longilinea	0.05	0.18	0.31	0.23	0.19	0.24	0.17	0.06	0.16	0.18	0.17	0.04	0.13	0.28	0.06	0.07	0.00	0.00	0.08	0.02	4.59	2.83
Other genera	42.72	46.06	49.62	42.57	53.21	54.65	44.16	59.85	59.78	59.42	63.15	75.38	64.91	54.84	55.85	57.89	70.41	70.24	67.03	71.70	46.94	51.11
Unclassified	10.01	9.34	8.34	9.65	14.05	20.01	8.83	10.58	9.55	10.41	12.11	6.62	11.78	15.05	10.37	15.62	7.62	6.60	6.51	5.57	6.67	8.30

Table 3.35: Top 15 genera that have relative abundance >1% for total 40 sediment samples after sequence numbers normalization (including page 159).

	RF1_a1	RF1_a2	RF1_b2	SG1_a1	SG1_a2	SG2_a1	SG2_a3	SG2_b1	SG2_b2	SG2_b3	SG3_a1	SG3_b1	SG6_a1	SG6_a2	SG5_a1	SG5_a2	SG5_b1	SG5_b2
Uncultured Anaerolineaceae	7.58	4.19	9.15	8.28	8.68	6.97	6.87	7.91	6.70	7.01	7.51	4.69	9.72	8.21	10.73	8.62	7.34	6.87
Uncultured Nitrosomonadaceae	1.58	5.63	3.12	1.88	2.31	1.17	0.23	0.54	0.74	0.17	2.15	6.00	3.45	3.62	1.48	2.58	2.95	2.95
Uncultured Nitrospiraceae	4.46	5.26	4.09	4.29	3.99	10.49	8.08	6.20	8.05	4.89	2.11	2.08	1.27	1.37	3.52	1.71	2.92	2.58
Nitrospira	1.27	3.39	2.75	0.64	1.37	0.50	0.37	1.94	0.37	0.23	0.47	1.91	7.31	5.80	9.19	5.23	0.87	2.65
Spirochaeta	2.35	1.07	2.04	2.41	1.68	3.02	2.41	2.15	1.84	1.61	1.27	0.70	0.47	0.74	0.91	0.91	1.07	0.70
Anaeromyxobacter	0.37	0.60	1.24	1.58	2.08	0.20	0.10	0.40	0.30	0.17	4.09	6.13	0.84	0.77	0.10	0.10	0.54	0.40
Dechloromonas	0.00	0.13	0.13	1.44	2.11	0.07	0.10	0.20	0.10	0.03	4.43	3.55	4.16	3.89	0.10	0.00	0.17	0.13
Uncultured Alcaligenaceae	0.30	0.20	1.04	0.50	1.81	0.17	0.17	0.54	0.10	0.13	6.24	8.31	1.27	2.15	0.77	0.74	1.88	0.94
Uncultured Comamonadaceae	0.27	0.37	0.20	0.34	0.94	0.23	0.03	0.13	0.10	0.17	0.74	1.01	0.94	1.61	1.81	1.74	2.48	1.88
Uncultured Rhodocyclaceae	0.50	0.54	0.34	0.67	1.11	0.44	0.13	0.40	0.60	0.00	3.92	1.98	1.91	2.61	0.20	0.40	0.91	0.67
Acidiferrobacter	0.27	1.07	0.23	0.03	0.10	0.20	0.27	0.34	0.13	0.00	0.07	0.13	0.50	0.64	2.72	1.74	1.41	1.71
Uncultured Hydrogenophilaceae	0.03	0.20	0.34	0.07	0.40	0.00	0.00	0.13	0.20	0.00	0.47	0.37	0.97	1.11	1.24	1.11	0.84	0.97
Uncultured Syntrophaceae	0.27	0.13	0.23	0.74	0.27	0.20	0.30	0.50	0.64	1.21	0.37	0.13	0.03	0.10	0.20	0.44	0.03	0.23
Leptolinea	0.30	0.13	0.54	0.23	0.60	0.13	0.27	0.13	0.23	0.23	1.21	0.27	0.23	0.37	0.20	0.03	0.10	0.07
Longilinea	0.10	0.03	0.03	0.27	0.13	0.07	0.07	0.10	0.03	0.13	0.40	0.03	0.27	0.10	0.00	0.00	0.07	0.03
Other genera	12.06	21.58	16.37	12.03	19.55	10.90	5.44	11.41	10.71	5.94	22.08	20.94	23.05	22.26	15.04	18.60	19.93	20.70
Unclassified	68.29	55.48	58.16	64.60	52.87	65.24	75.16	66.98	69.16	78.08	42.47	41.77	43.61	44.65	51.79	56.05	56.49	56.52

	SG4_ a1	SG4_ b3	SG4_ b2	SG4_ b1	DN3_ a1	DN3_ a2	DN3_ b1	DN2_ a1	DN2_ a2	DN2_ a3	DN2_ b1	DN2_ b2	DN2_ b3	DN1_ a1	DN1_ a2	DN1_ b1	RF2_ a1	RF2_ a2	RF2_ b1	RF2_ b2	SG8_ a1	SG9_ a1
Uncultured Anaerolineaceae	9.45	10.59	7.91	10.49	9.25	7.51	8.82	7.44	8.05	6.50	7.21	3.82	5.26	10.22	7.81	8.72	5.33	4.46	5.53	4.56	14.95	13.85
Uncultured Nitrosomonadaceae	2.08	1.17	0.97	1.61	0.67	0.03	2.08	0.94	1.21	1.54	2.28	2.51	2.18	0.54	2.58	0.84	3.49	3.62	3.62	3.25	0.20	0.07
Uncultured Nitrospiraceae	1.81	4.22	1.88	2.51	1.88	1.98	3.05	1.61	2.51	1.94	2.45	1.14	1.37	1.74	4.09	2.08	0.50	0.74	0.30	0.80	0.20	0.30
Nitrospira	12.67	9.29	10.33	9.92	2.75	0.07	12.20	0.64	1.07	0.80	1.34	1.84	1.91	1.17	4.79	0.97	2.08	2.04	3.18	2.28	0.03	0.13
Spirochaeta	0.57	1.34	0.91	0.97	2.61	2.78	0.94	1.58	0.91	1.88	0.80	0.37	1.21	2.01	0.97	2.68	0.40	0.67	0.94	0.57	0.30	0.20
Anaeromyxobacter	0.20	0.13	0.34	0.13	0.07	0.00	0.23	1.27	1.14	1.58	0.67	0.87	1.17	0.47	0.77	0.54	1.34	1.51	1.84	1.78	0.07	0.03
Dechloromonas	0.00	0.17	0.37	0.03	0.07	0.00	0.27	1.84	2.28	2.18	0.40	0.67	0.70	0.94	0.47	0.10	0.67	0.91	0.97	1.07	4.09	0.87
Uncultured Alcaligenaceae	0.74	0.67	0.80	1.11	0.13	0.00	0.70	1.84	1.21	1.04	0.94	0.54	1.17	0.50	0.94	0.30	0.37	0.60	0.20	0.64	0.10	0.47
Uncultured Comamonadaceae	1.94	2.28	1.41	1.94	0.94	0.03	2.68	0.64	0.94	0.77	0.67	1.58	0.60	0.30	0.50	0.30	1.48	1.58	1.48	2.04	0.80	0.23
Uncultured Rhodocyclaceae	0.34	1.17	1.07	0.60	0.27	0.03	0.47	2.18	1.81	2.08	0.37	0.80	0.50	1.01	0.94	0.23	0.94	1.17	0.97	0.77	2.04	0.60
Acidiferrobacter	1.81	2.18	1.68	2.28	0.64	0.03	2.58	0.13	0.03	0.03	0.13	0.00	0.03	0.07	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Uncultured Hydrogenophilaceae	0.94	1.24	1.07	0.87	0.23	0.00	1.07	0.10	0.20	0.27	0.10	0.00	0.10	0.20	0.10	0.00	0.30	0.13	0.27	0.00	0.23	1.21
Uncultured Syntrophaceae	0.64	0.54	0.47	0.50	1.24	1.48	0.37	0.74	0.87	1.17	0.30	0.27	0.70	1.21	0.87	1.64	0.03	0.27	0.17	0.10	0.57	1.41
Leptolinea	0.07	0.47	0.47	0.07	0.07	0.23	0.17	0.20	0.37	0.34	0.17	0.23	0.10	0.60	0.10	0.30	0.17	0.10	0.20	0.03	11.50	7.04
Longilinea	0.07	0.20	0.23	0.17	0.17	0.37	0.10	0.03	0.23	0.30	0.07	0.07	0.10	0.40	0.03	0.07	0.03	0.03	0.13	0.03	4.02	2.25
Other genera	14.54	14.36	18.73	16.58	10.79	6.98	18.21	23.17	23.06	23.44	23.84	40.77	28.22	13.75	20.39	17.50	35.33	37.75	34.14	39.07	28.38	21.09
Unclassified	52.13	49.98	51.36	50.22	68.22	78.48	46.06	55.65	54.11	54.14	58.26	44.52	54.68	64.87	54.48	63.73	47.54	44.42	46.06	43.01	32.52	50.22

3.5. Searching for correlations:

3.5.1. Searching for the geographic correlation:

3.5.1.1. Principle component analysis (PCA) at the phyla level:

3.5.1.1.1. PCA with PC1 & PC2, PC1 & PC3, PC2 & PC3 of 40 samples:

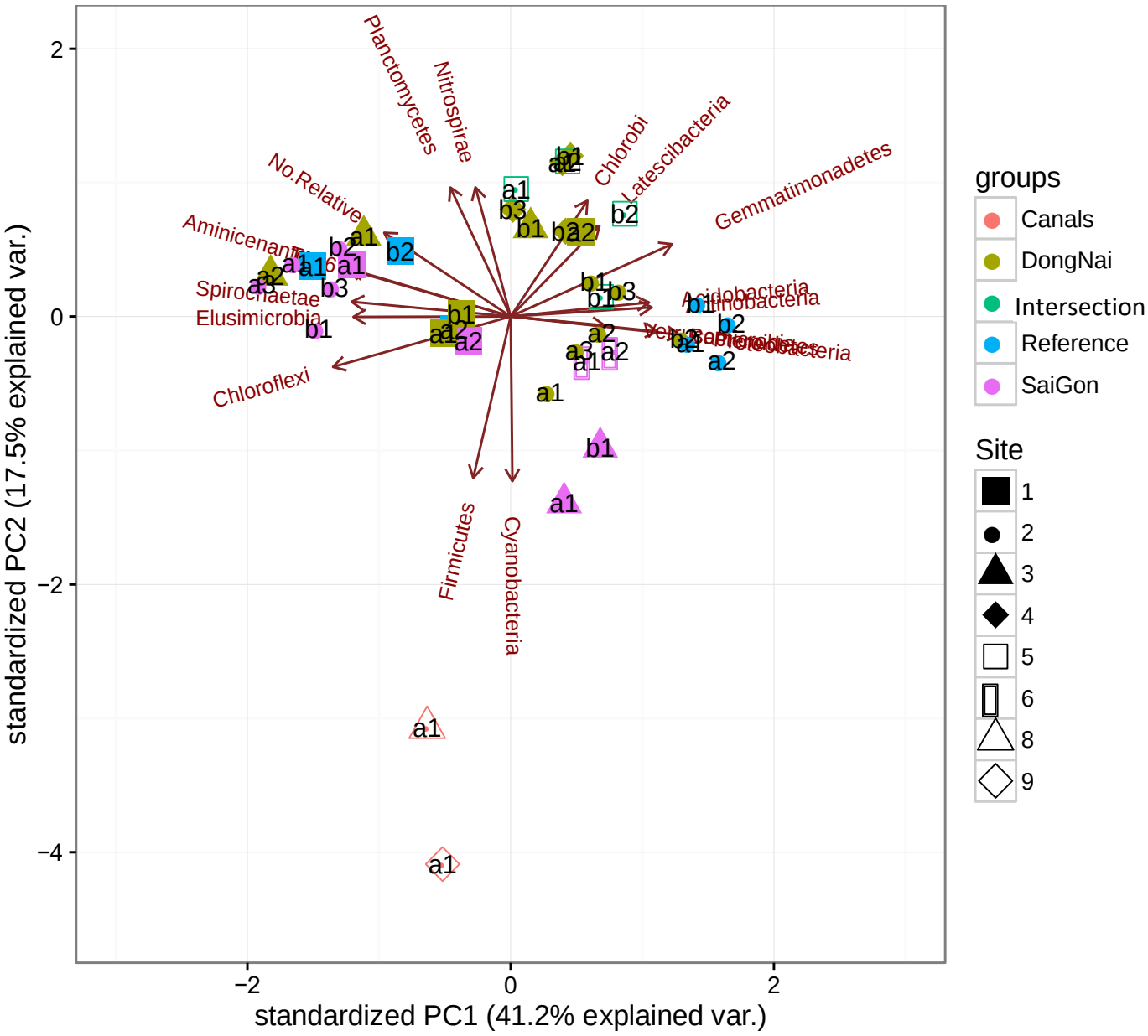


Figure 3.29. PCA GG plot of the bacterial communities based on phyla of 40 samples from the SG-DN river system on the first two principal components (PC1 & PC2). The most 18 abundant phyla were selected for the analysis.

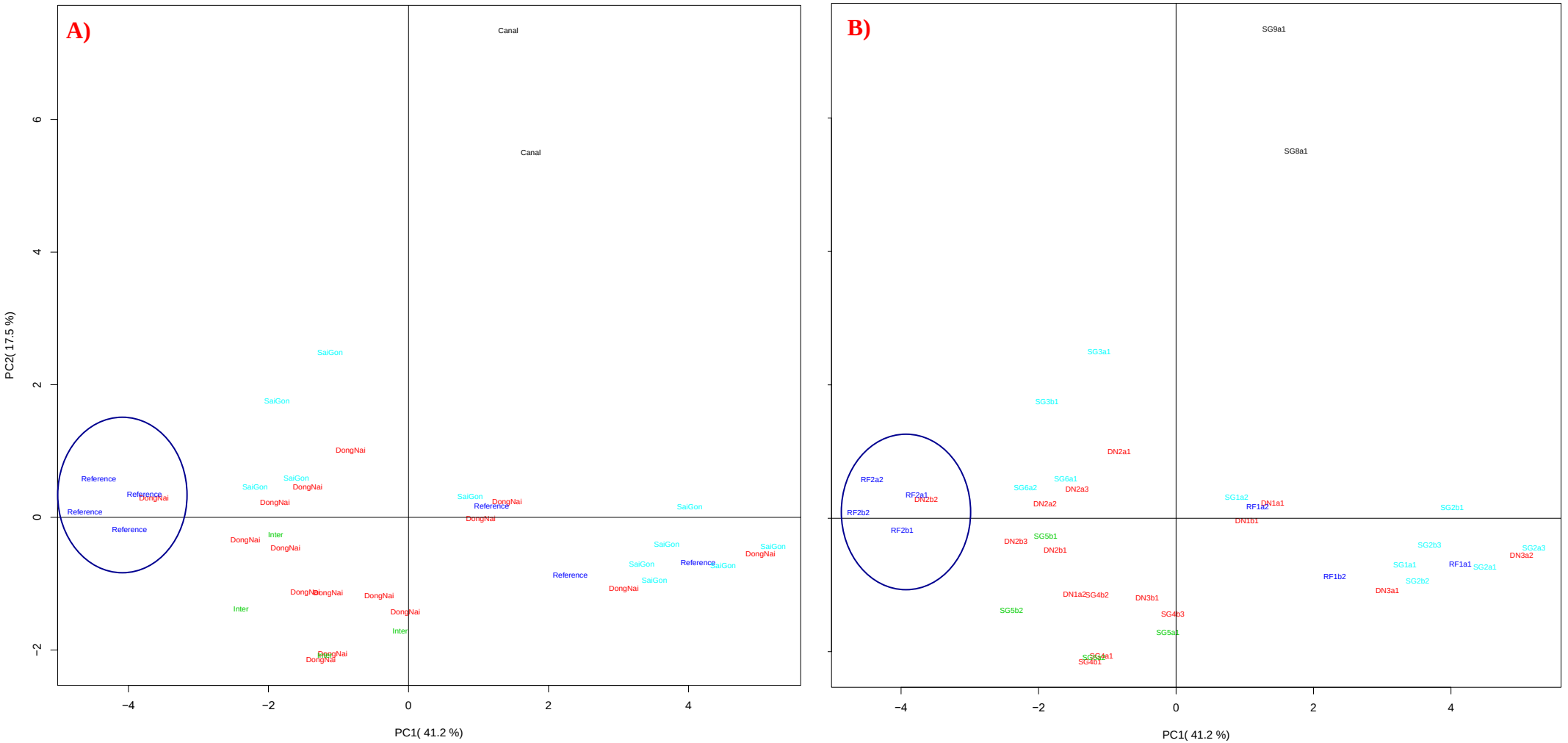


Figure 3.30. PCA CC plot of the bacterial communities based on phyla of 40 samples from the SG-DN river system on the first two principal components (PC1 & PC2). The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

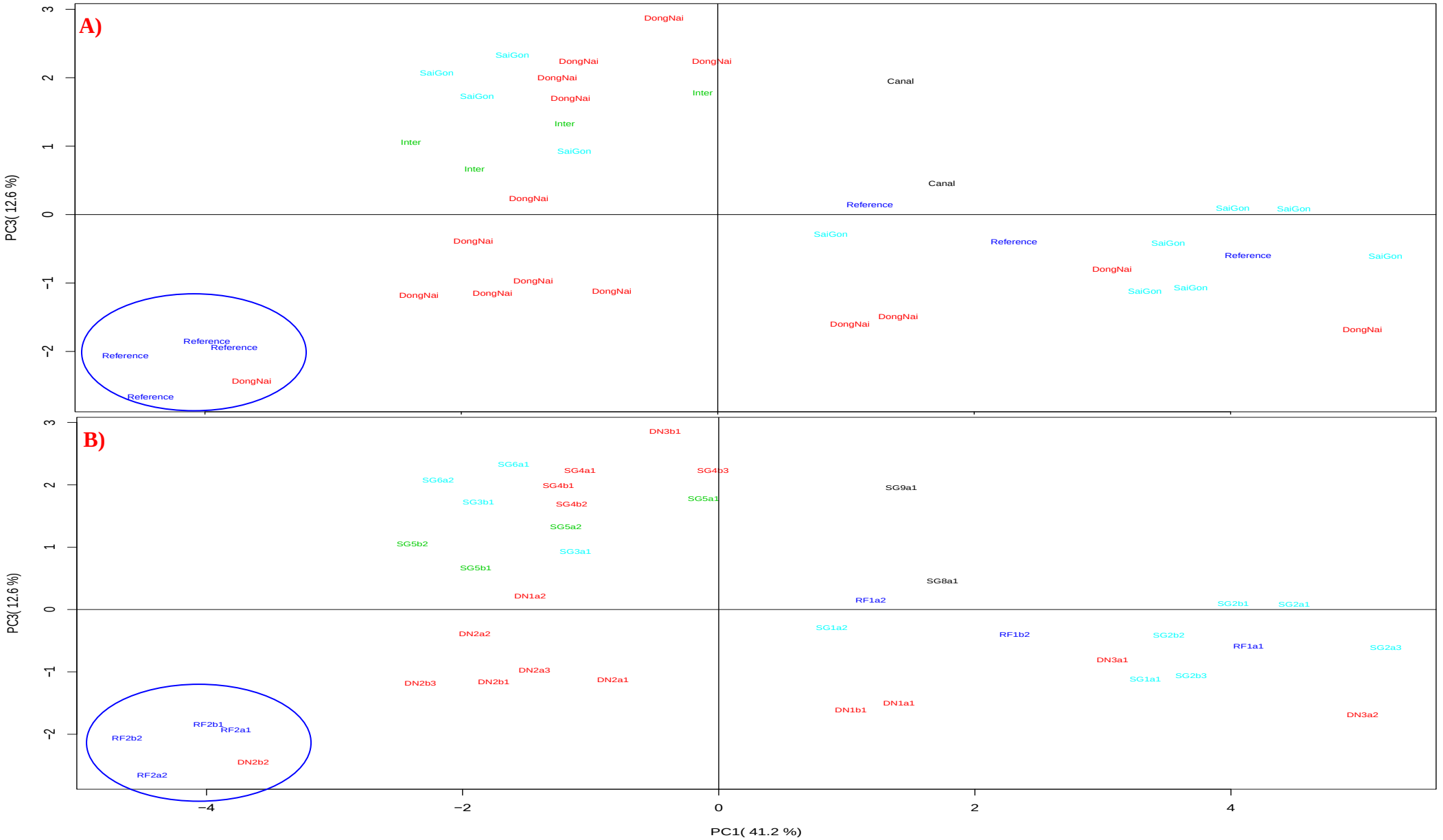


Figure 3.31. PCA CC plot of the bacterial communities based on phyla of 40 samples from the SG-DN river system on the first and third principal components (PC1 & PC3). The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

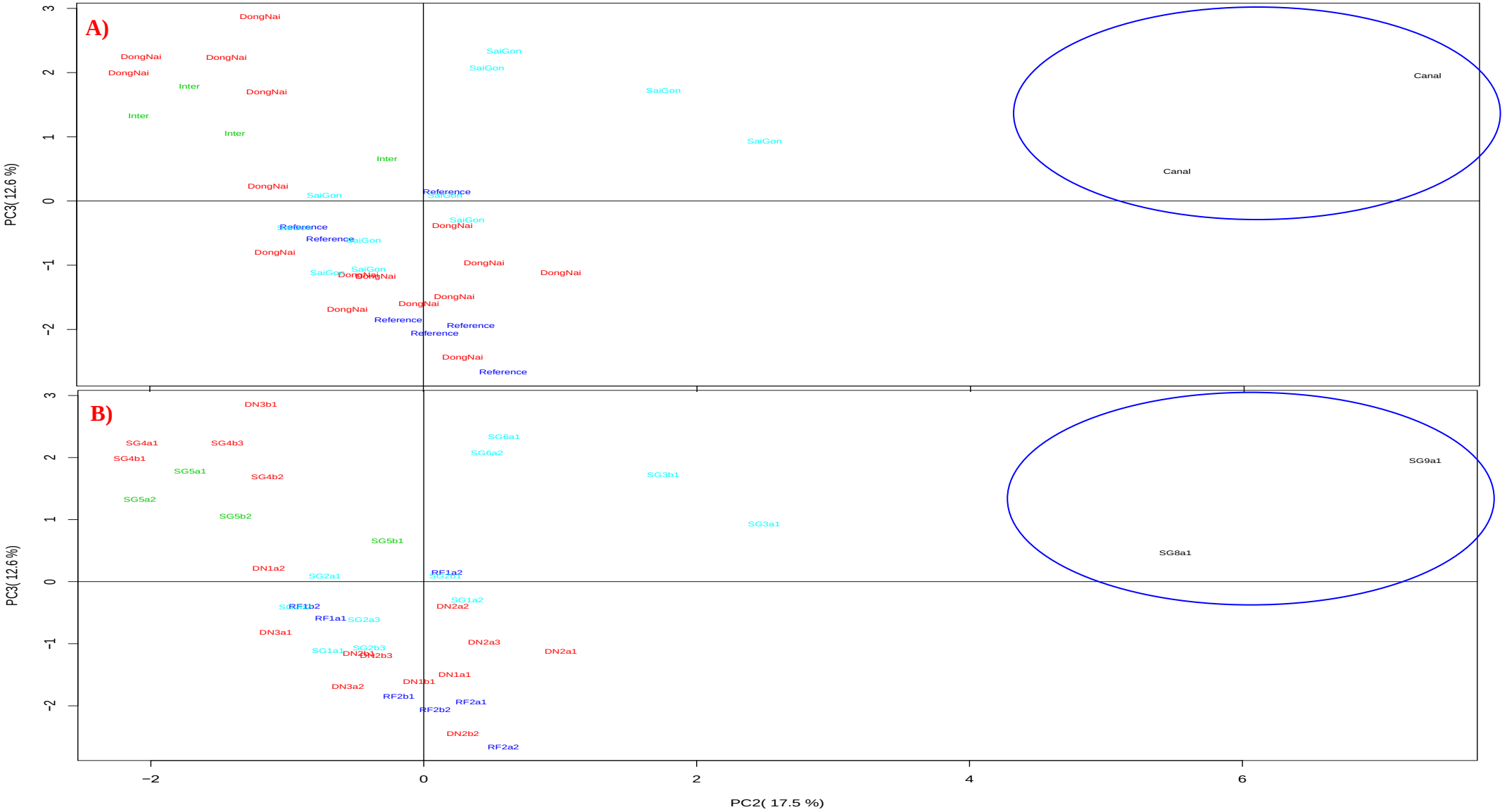


Figure 3.32. PCA CC plot of the bacterial communities based on phyla of 40 samples from the SG-DN river system on the second and third principal components (PC2 & PC3). Top 18 phyla were selected for the analysis. The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

3.5.1.1.2. PCA with PC1 & PC2, PC1 & PC3, PC2 & PC3 of 38 samples (without SG8a1 & SG9a1):

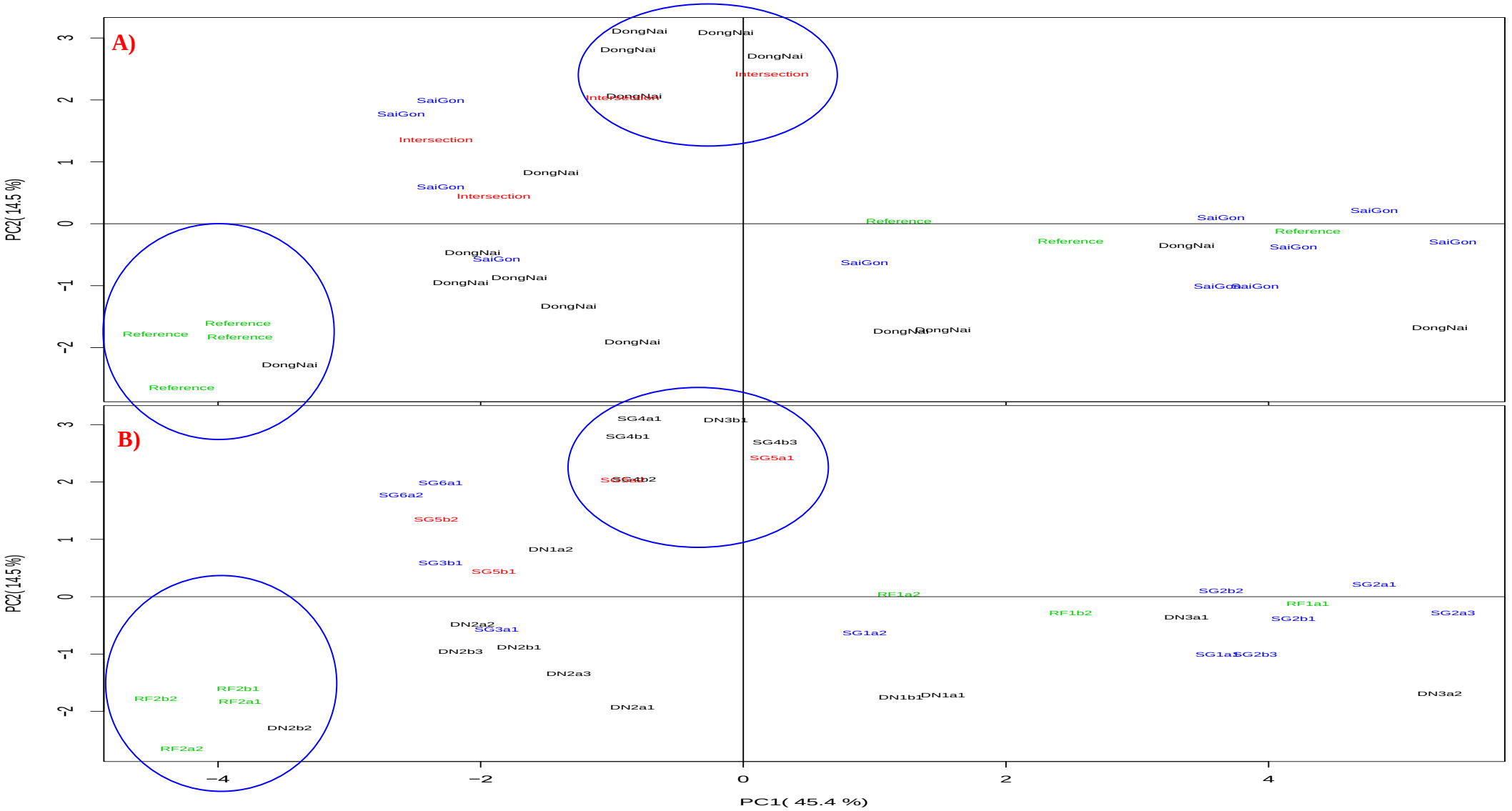


Figure 3.33. PCA CC plot of the bacterial communities based on phyla of 38 (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the first two principal components (PC1 & PC2). The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

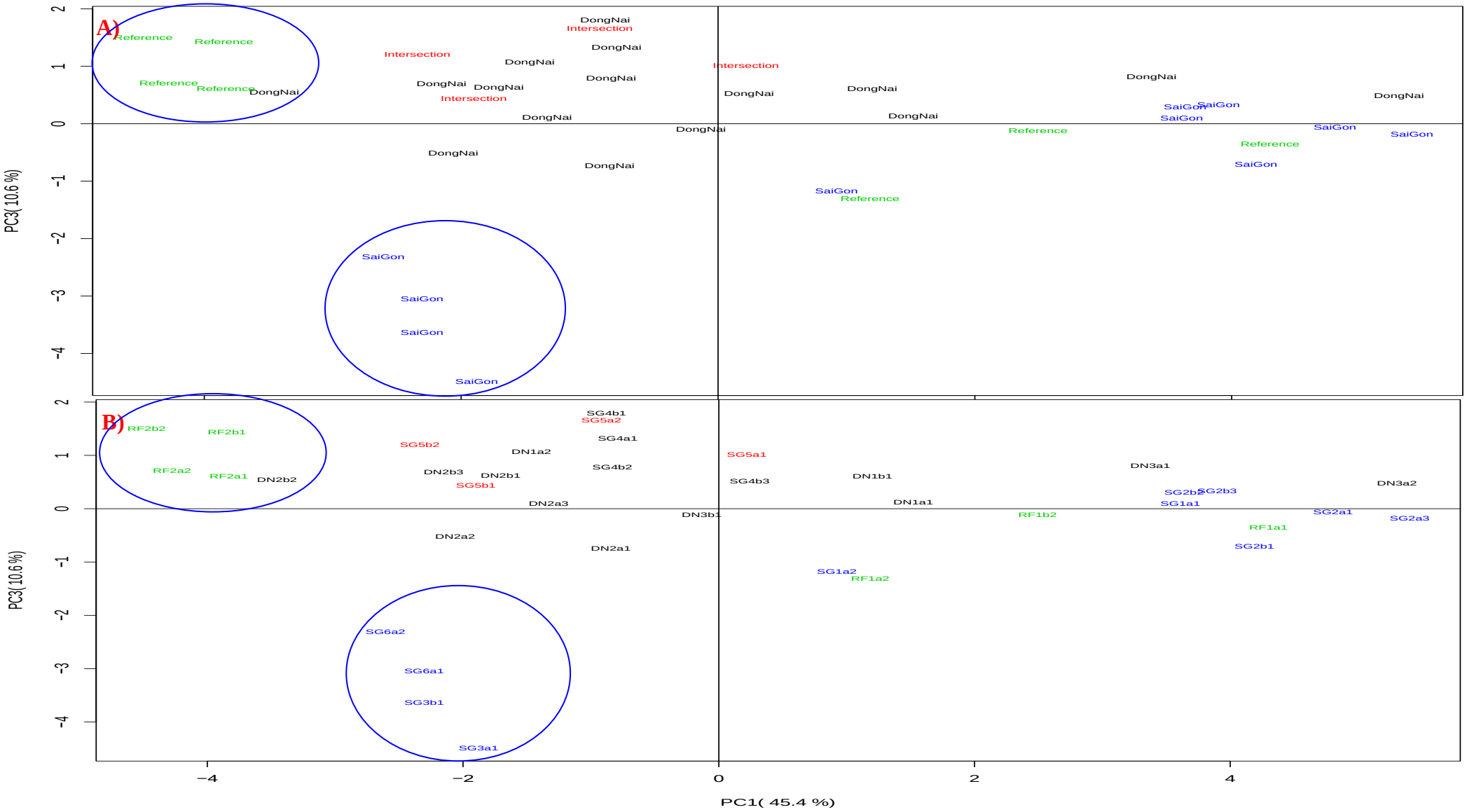


Figure 3.34. PCA CC plot of the bacterial communities based on phyla of 38 (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the first and third principal components (PC1 & PC3). The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

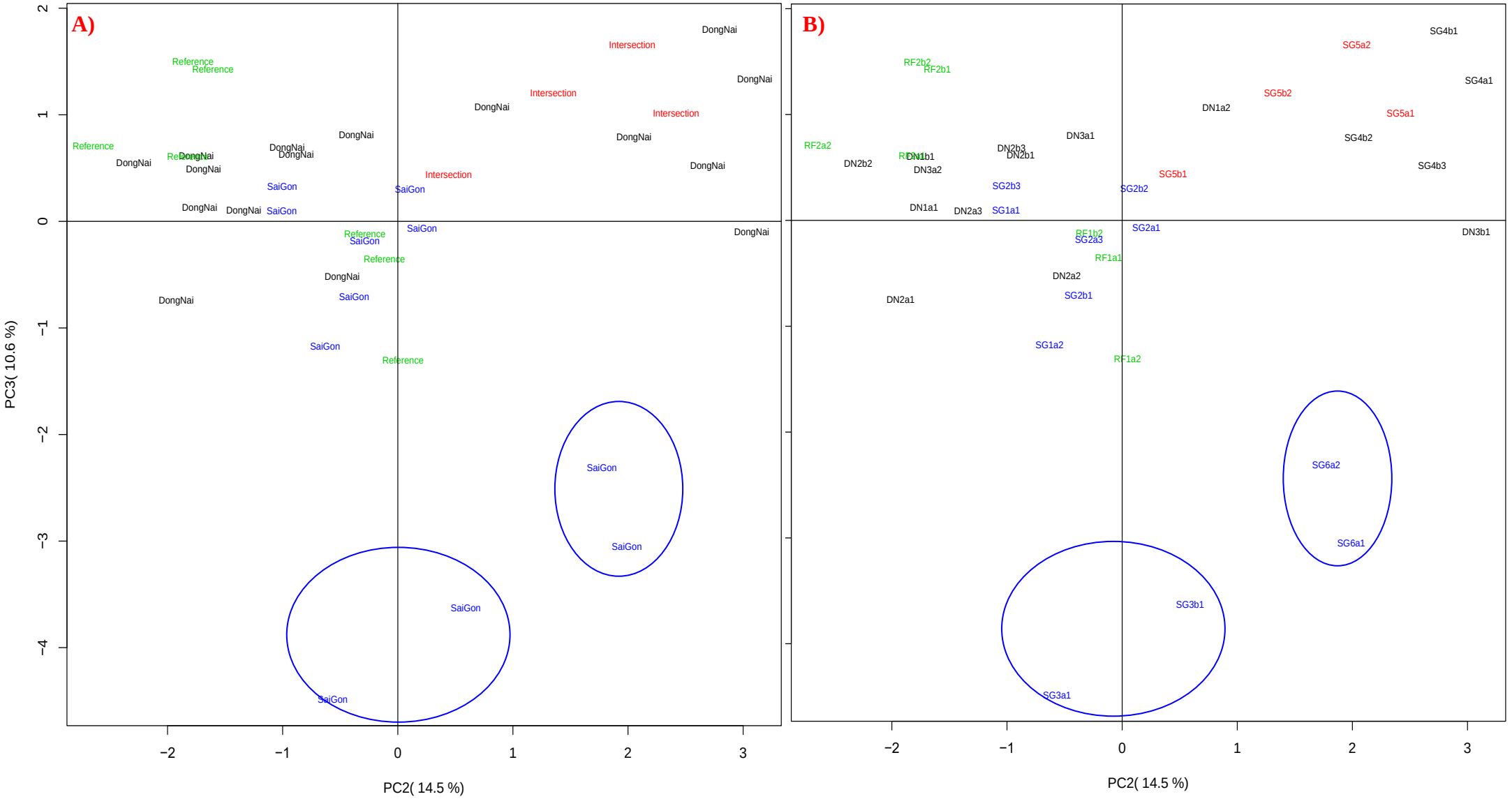


Figure 3.35. PCA CC plot of the bacterial communities based on phyla of 38 (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the second and third principal components (PC2 & PC3). The most abundant 18 phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

In order to investigate the different components could affect the grouping of the samples according to 18 most dominant phyla, different PCA analyses in CC plot with PC1 & PC2, PC2 & PC3 and PC1 & PC3 were performed for i) 40 samples with canal samples SG8a1 & SG9a1, ii) 38 samples without SG8a1 & SG9a1 (**Fig. 3.22-30**, subsequently).

i) *40 samples with canal samples SG8a1 & SG9a1:*

Three Principle Component (PC) analyses which were performed in order to access the clustering and the behavior of sediment samples from the SG-DN river system via the most 18 abundant phyla. PC1 & PC2 and PC2 & PC3, which explain 58.7% and 30.1% of total variance, respectively, separated samples SG8a1 & SG9a1 from the rest of the samples (**Fig.3. 30**).

In GG plot, the first two axes of the PCA (**Fig. 3.29**), which explained 58.7% of the total variance where the samples were found to cluster by collection into different groups and influenced principally by the proportions of several particular phyla members. the phyla *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia* and *Proteobacteria* oriented the grouping of RF2 (a1, a2, b1, b2) and DN2b2, which can be explained by the fact that samples RF2 (a1, a2, b1, b2) and DN2b2 had highest proportions of *Acidobacteria* among the samples (from 10.1% to 7.6%) (**Fig. 3.29 &Table 3.33**). However, the abundance of *Actinobacteria* of samples RF2 (a1, a2, b1, b2) are highest among the samples (from 2.9% to 4.7%) but not DN2b2 (1.4%). The abundance of *Verrucomicrobia* of sample DN2b2 was highest among the samples (4.0 %), then RF2a2 (3.5%) and those of samples RF2 (a1, b1, b2) were quite high among the samples (from 1.2% to 1.7%).

Samples SG8a1 & SG9a1 formed the outgroup due to their highest proportion of both *Cyanobacteria* (3.6%, 16.0%) and *Firmicutes* (2.6%, 2.9%) among the samples. This result is agreement with the cluster in UPGMA tree (**Fig. 3.43**). The clustering of samples RF2 (a1, a2, b1, b2) and DN2b2 are properly due to the high proportion of *Bacteroidetes* (4.6%-6.2%) among 40 samples. However for *Verrucomicrobia*, only samples RF2a2 and DN2b2 had the highest proportion (3.5% and 4.0%, respectively).

For the *Proteobacteria*, these samples had similar proportions (from 35.4%-41.1%), as seen with *Acidobacteria* (from 8.4% - 10.1%). For *Actinobacteria*, highest proportion belong to samples RF2 (a1, a2, b1, b2) (2.9%-4.7%) but not DN2b2 (1.4%). In summary, the high value of three main phyla (*Bacteroidetes*, *Proteobacteria*, *Acidobacteria)* oriented the cluster of 5 sites RF2 (a1, a2, b1, b2) and DN2b2, which belong to the DongNai river.

ii) 38 samples without SG8a1 & SG9a1:

PC1 & PC3 which explained of 56.0% total variance showed the separation between the SaiGon and the DongNai rivers (**Fig.3. 33-35**).

3.5.1.2. Principle component analysis (PCA) at genus level:

3.5.1.2.1. PCA with PC1 & PC2, PC1 & PC3, PC2 & PC3 of 40 samples:

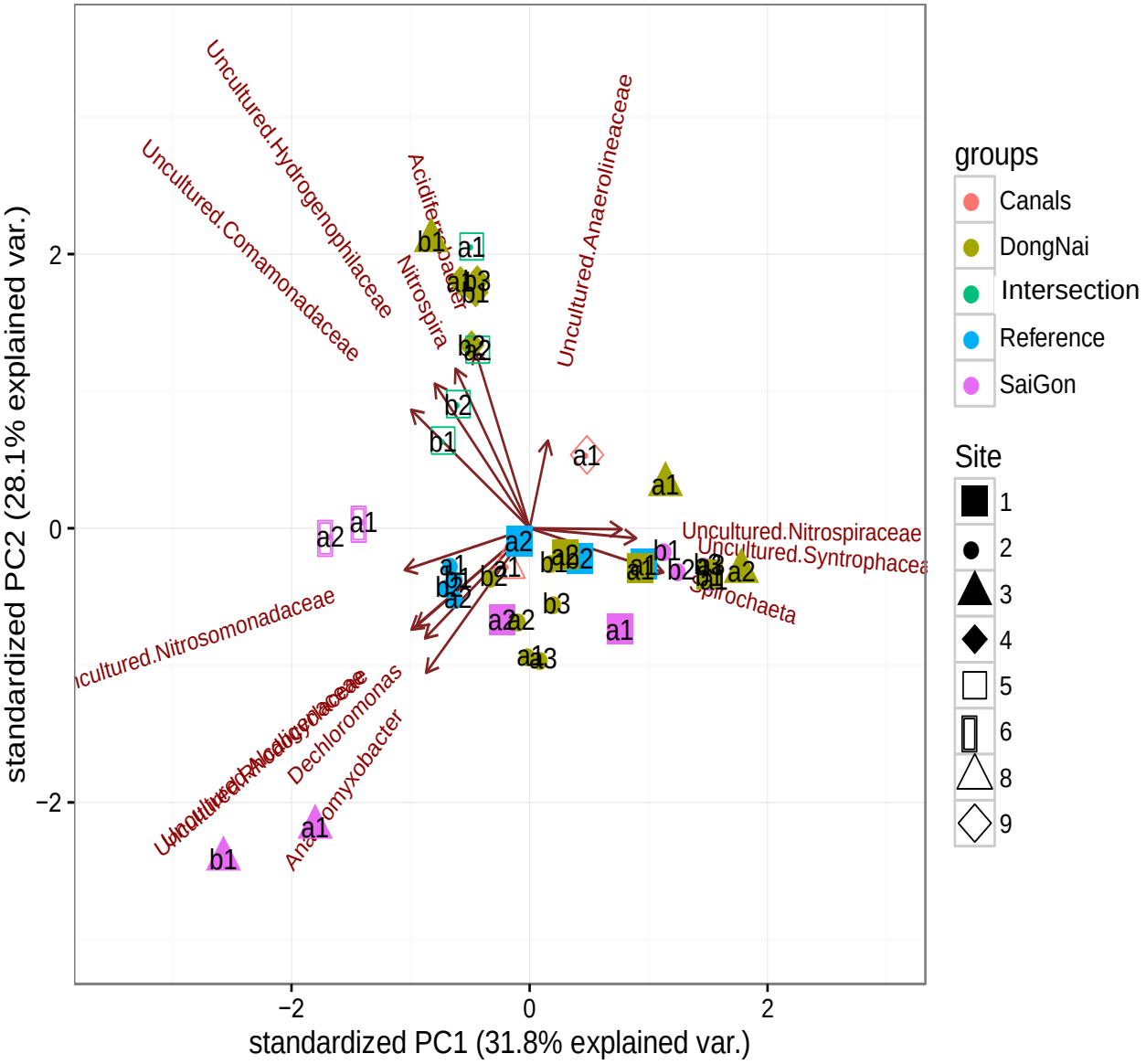


Figure 3.36. PCA GG plot PC1 & PC2 of the bacterial communities based on genera of 40 sediment samples from the SG-DN river system on the first two principal components. The most 13 abundant phyla were selected for the analysis.

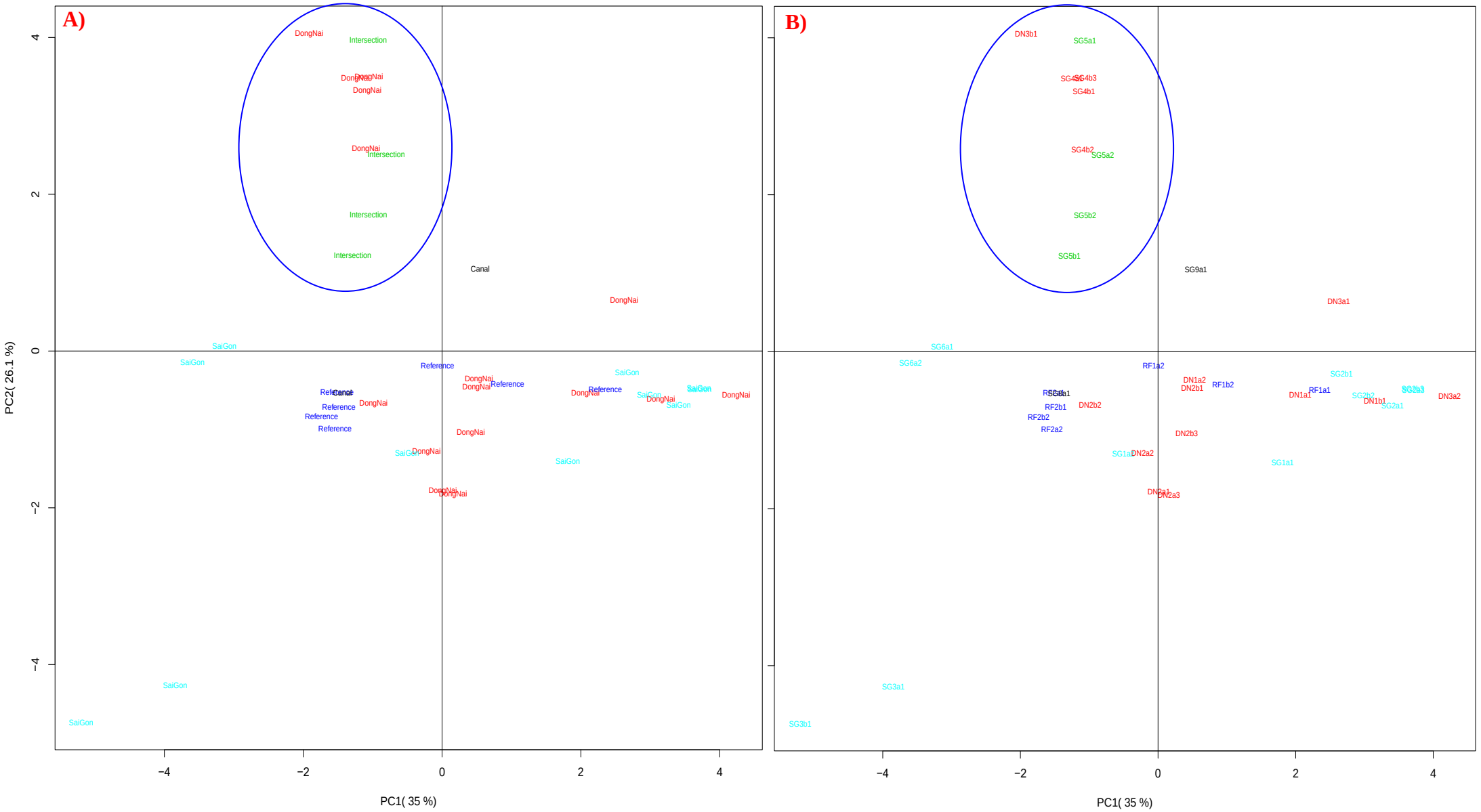


Figure 3.37. PCA CC plot of the bacterial communities based on genera of 40 sediment samples from the SG-DN river system on the first two principal components (PC1 & PC2). The most 13 abundant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

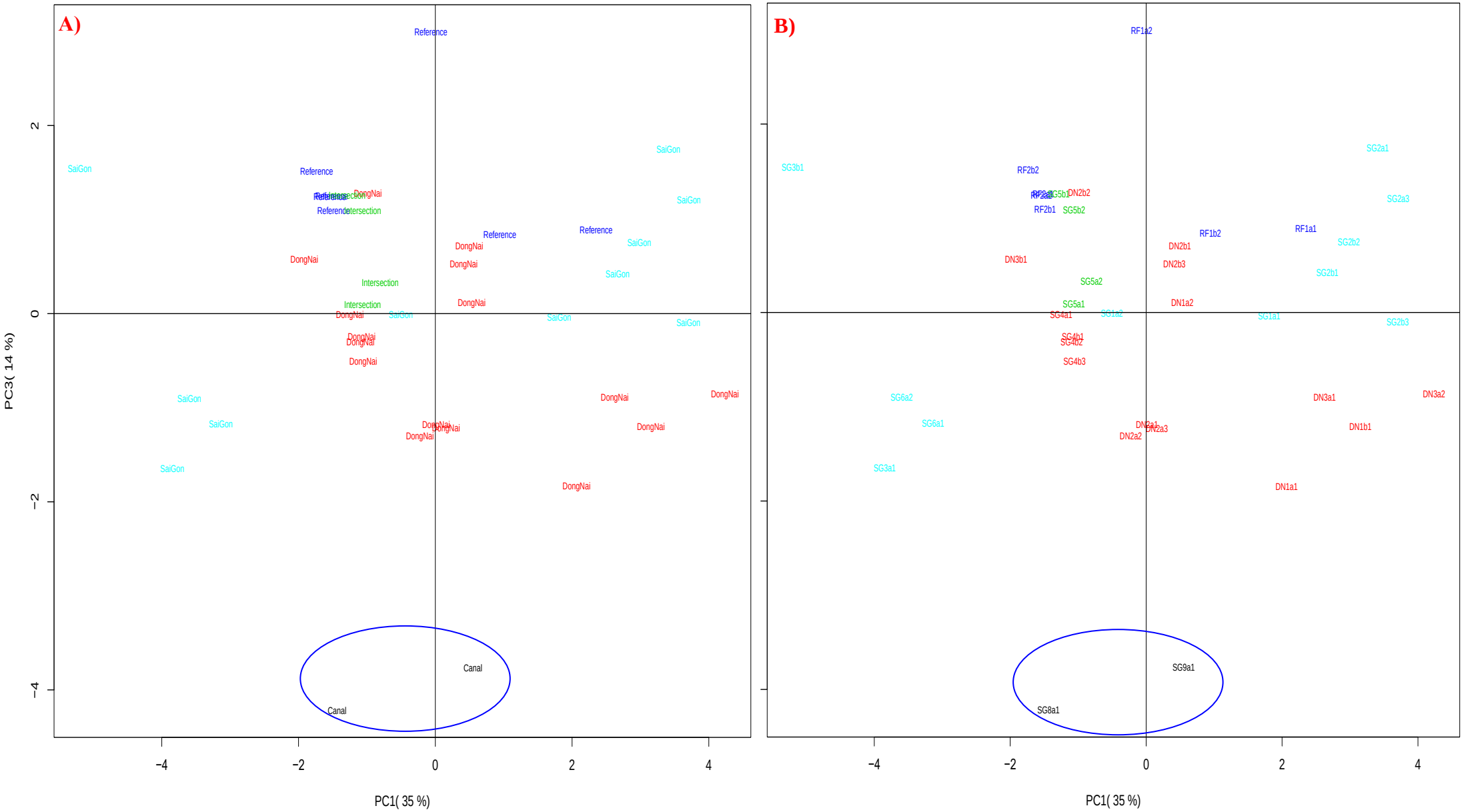


Figure 3.38. CC plot of the bacterial communities based on genera of 40 sediment samples from the SG-DN river system on the first and third principal components (PC1 & PC3). The most 13 abundant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

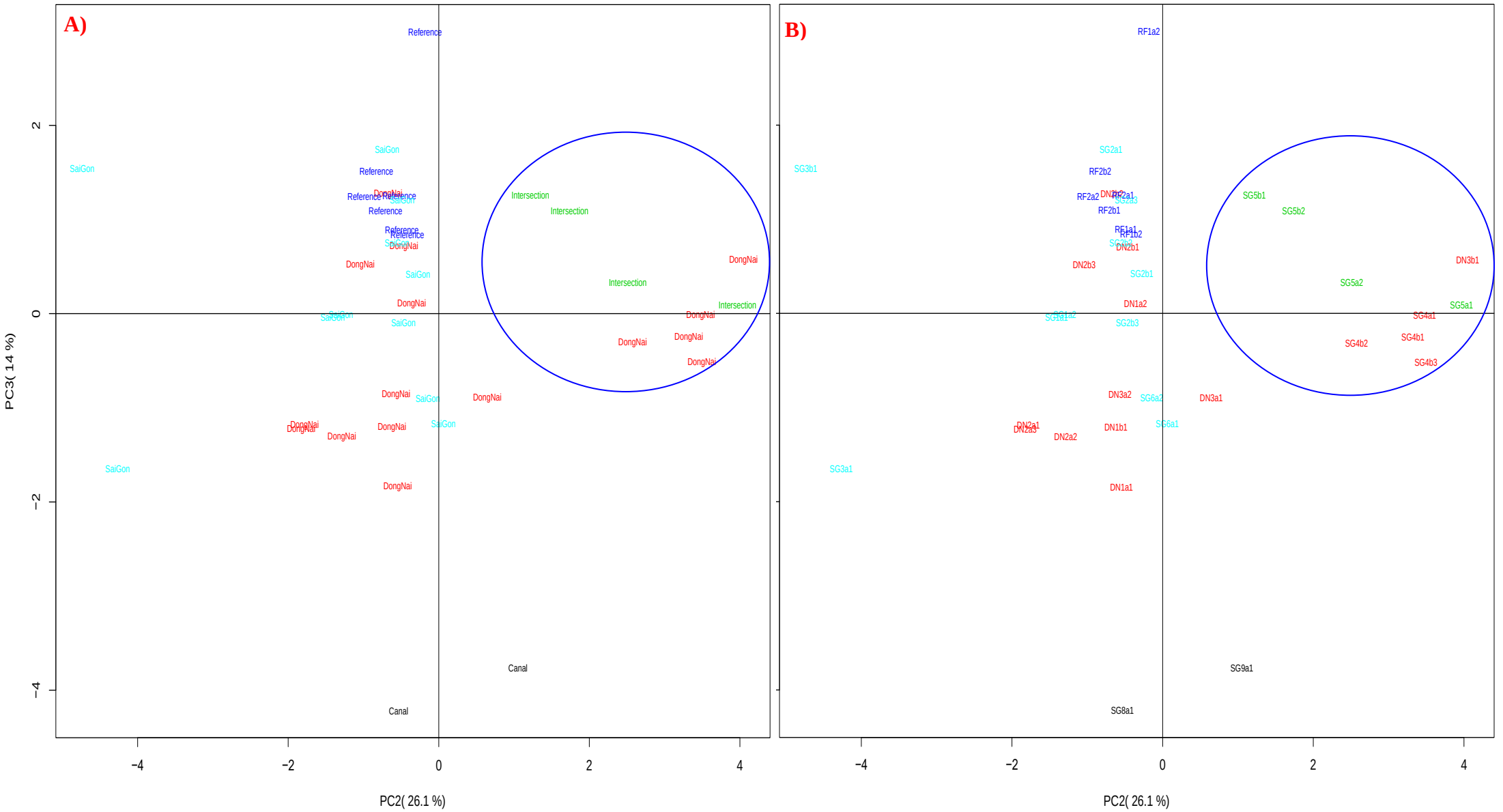


Figure 3.39. CC plot of the bacterial communities based on genera of 40 sediment samples from the SG-DN river system on the second and third principal components (PC2 & PC3). The most 13 abudant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

3.5.1.2.2. PCA with PC1 & PC2, PC1 & PC3, PC2 & PC3 of 38 samples (without SG8a1 & SG9a1):

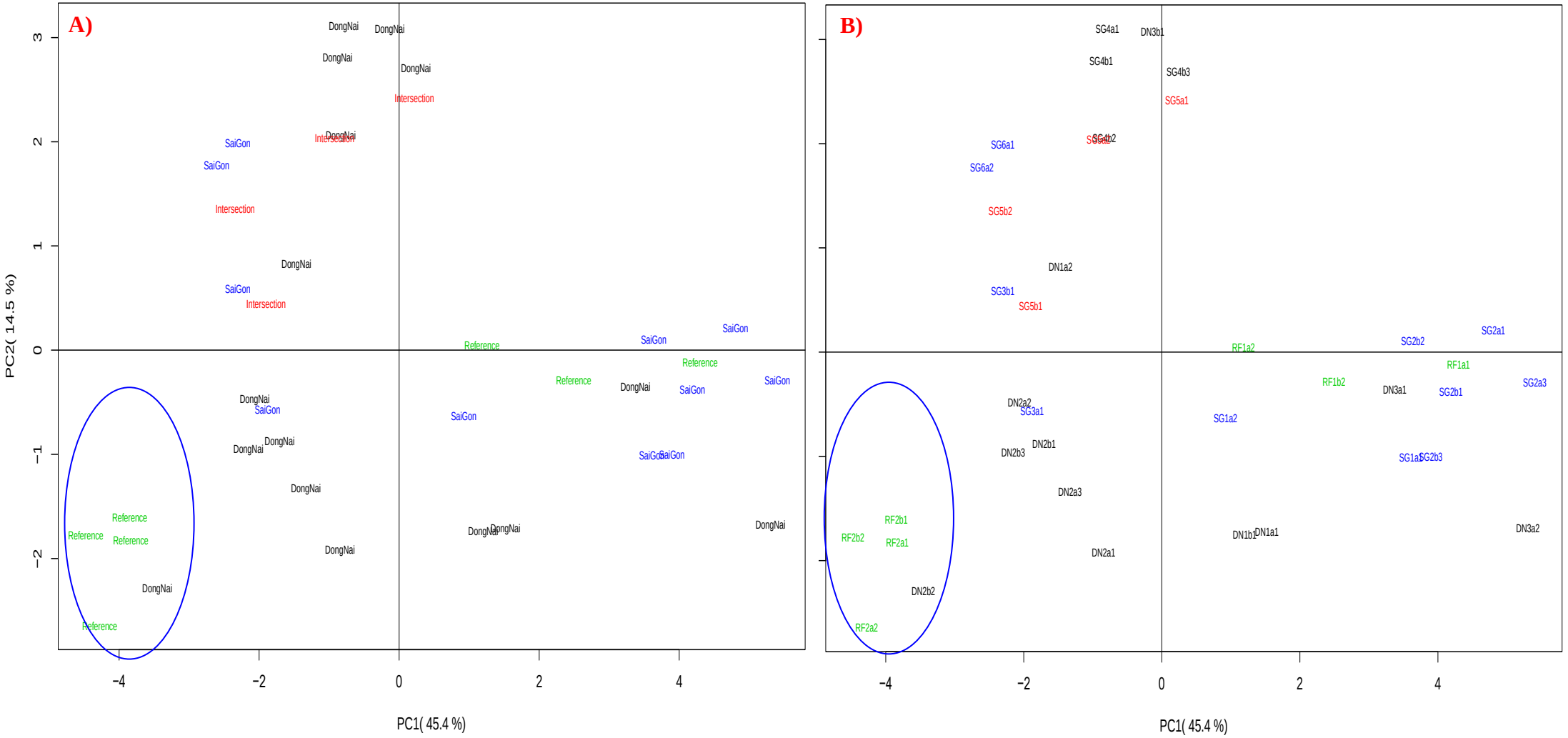


Figure 3.40. CC plot of the bacterial communities based on genera of 38 samples (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the two first components (PC1 & PC2). The most 13 abundant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

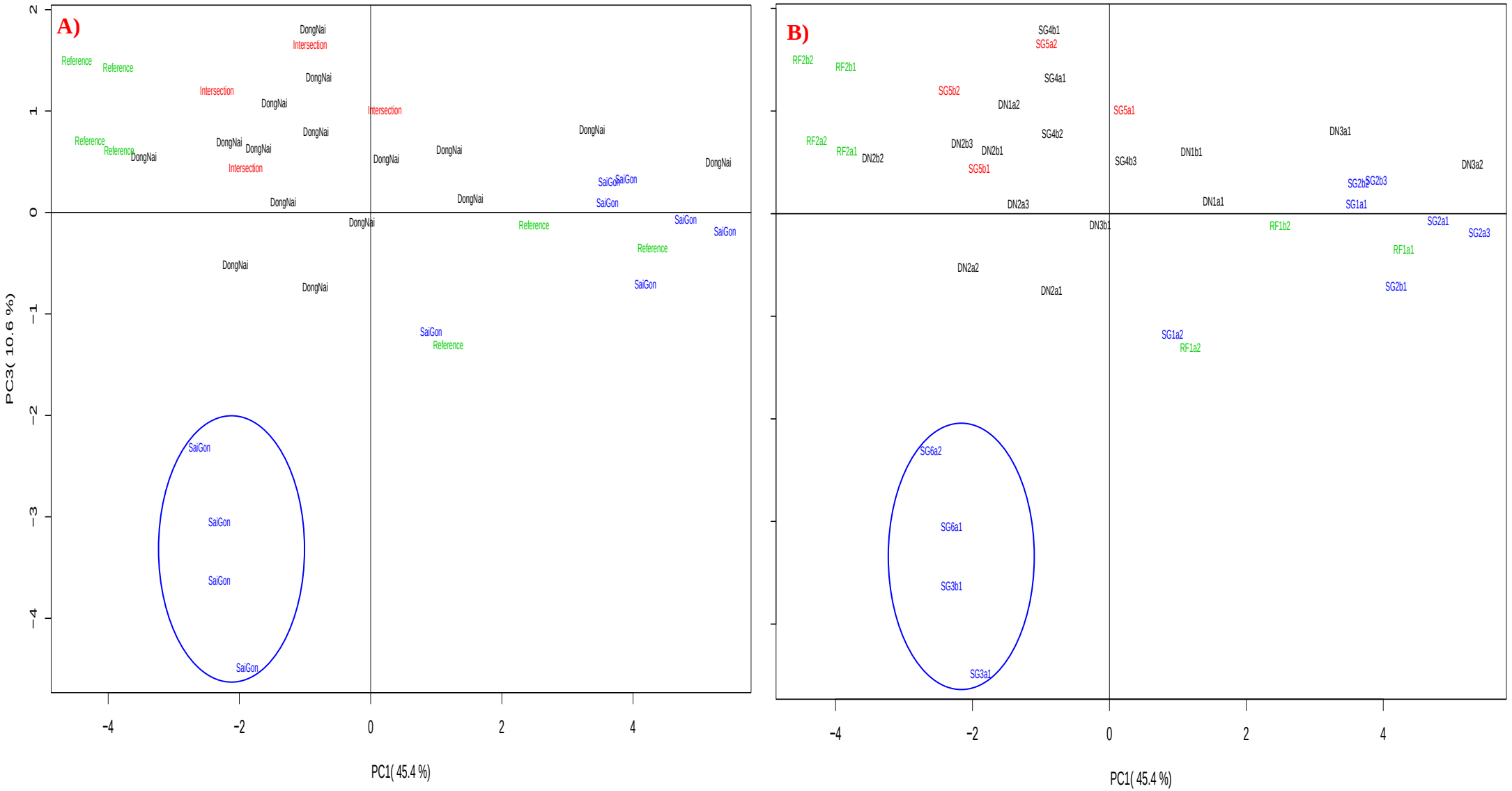


Figure 3.41. CC plot of the bacterial communities based on genera of 38 samples (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the first and third components (PC1 & PC3). The most 13 abundant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

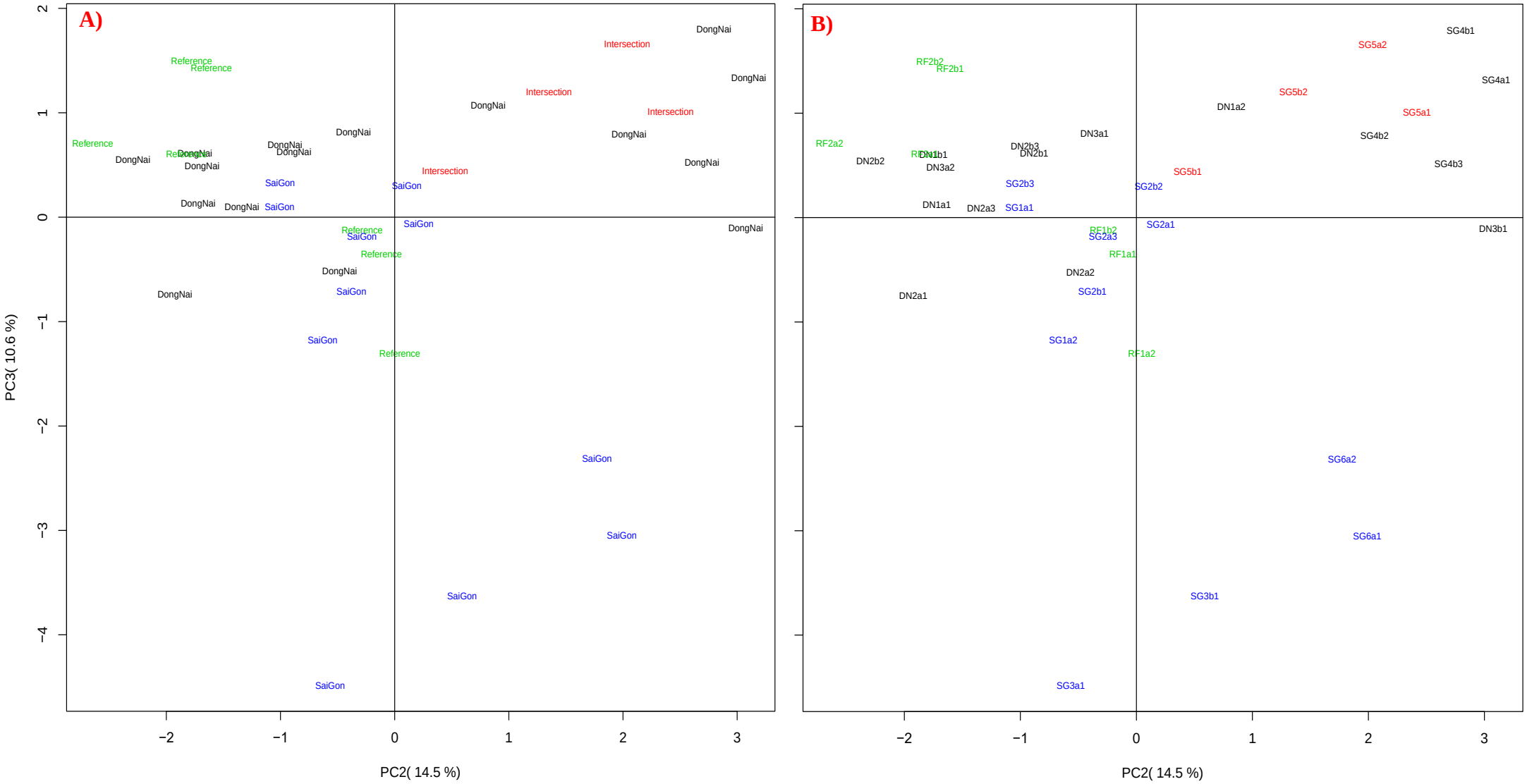


Figure 3.42. CC plot of the bacterial communities based on genera of 38 samples (without samples SG8a1 & SG9a1) samples from the SG-DN river system on the second and third components (PC2 & PC3). The most 13 abundant phyla were selected for the analysis. Note: **A)** Samples were organized according the location groups, which are the References (locations RF1 and RF2), SaiGon river, DongNai river, Intersection and Canals. **B)** Samples were organized by their names.

PCA with PC1 (35%) does not separate the samples from the SaiGon and the DongNai rivers. Noticeably, PC1 & PC2, which explains 61.1% of the variance, distinguish samples SG3a1 and SG3b1 from the rest. In contrast, PC1 & PC3, which explains 49% of the variance, separate samples SG8a1& SG9a1 from the others, as well as samples SG6a1 & SG6a2. In PC2 & PC3, which explains 40.1% of the variance, separate samples SG3a1 & SG3b1, and sample RF1a2 from the others.

High abundance of genera *Anaeromyxobacter, Dechloromonas, Uncultured Alcaligenaceae* directed the grouping of samples SG3a1 & SG3b1 in GG plot (**Fig. 36**). Samples SG3a1 & SG3b1 had highest abundance of *Anaeromyxobacter, Dechloromonas* and *Uncultured Alcaligenaceae* among the samples (4.09%, 6.13%; 4.43%, 3.55%; 6.24%, 8.31%, respectively). *Uncultured Rhodocyclaceae* abundance is also highest in sample SG3a1 (3.92%) but not in sample SG3b1 (1.98%)

Similarity, high abundance of general *Dechloromonas* and *Uncultured Nitrosomonadaceae* oriented the grouping of samples SG6a1 & SG6a2. Samples SG6a1 & SG6a2 had high proportion of *Dechloromonas* (4.16% & 3.89%) and average abundance of *Uncultured Nitrosomonadaceae* (3.45% & 3.62%) among 40 samples. High abundance of genera *Uncultured Anaerolineaceae* distinguished samples SG8a1 & SG9a1. Samples SG8a1 & SG9a1 had highest abundance of *Uncultured Anaerolineaceae* (14.95% & 13.85%). *Dechloromonas* appears in the highest abundance in the samples SG3 (a1, b1), SG6 (a1, a2).

It appeared that there is an accumulation of *Nitrospira* in the downstream of the SaiGon and the DongNai rivers including the ontersection location. Interestingly, phylum *Nitrospira* also presented with a dominant abundance in the same samples, including SG5a1, SG4 (a1, a2, a3, b1) and DN3b1. There is an accumulation **of** phylum *Nitrospira* and genus *Nitrospira* in the downstream of the SaiGon and the DongNai rivers, including the intersection location.

Acidiferrobacter and *Uncultured Hydrogenophilaceae* abundance had significant proportion in samples of Intersection locaiton, SG5 (a1, a2, b1, b2), samples downstream of DongNai river SG4 (a1, b1, b2, b3) and DN3b1, indicating that there was also an accumulation of *Acidiferrobacter* and *Uncultured Hydrogenophilaceae* downstream of the SG-DN river system, including the intersection location.

Acidobacteriaceae (Subgroup 1) abundance were highest in the samples belonging the upstream of DongNai river, samples RF2 (a1, a2, b1 b2) with noticcably proportion ranging from 1.78% to 2.35%. Interestingly, phyla *Acidobacteria* also presented

with a dominant abundance in the same samples RF2 (a1, a2, b1 b2) with proportion from 8.4% to 10.1%.

Samples of canal locations, including SG8a1 & SG9a1 had highest abundance of *Uncultured Anaerolineaceae* (14.95% & 13.85%), *Leptolinea* (11.50% & 7.04%) and *Longilinea* (4.02% & 2.25%) among all the samples, indicating special ecological role of these genera in canal ecosystem in HCMC. Samples SG8a1 had significant *Uncultured Rhodocyclaceae* proportion (2.04%), *Dechloromonas* (4.09%), which may be specialized indicator for location SG8. Similarity, *Uncultured Syntrophaceae* and *Uncultured Hydrogenophilaceae* appreared significant proportion in sample SG9a1, specializing this location from SG8 location.

Samples of location SG3, including SG3a1 & SG3b1 had highest abundance of genera *Anaeromyxobacter* and *Uncultured Alcaligenaceae* among 40 samples with the proportion form 4.09%-6.13% & 6.24% -8.31%, respectively, indicating the esspecially ecological characteristics of this location.

3.5.1.3. UPGMA:

3.5.1.3.1. At the phyla level:

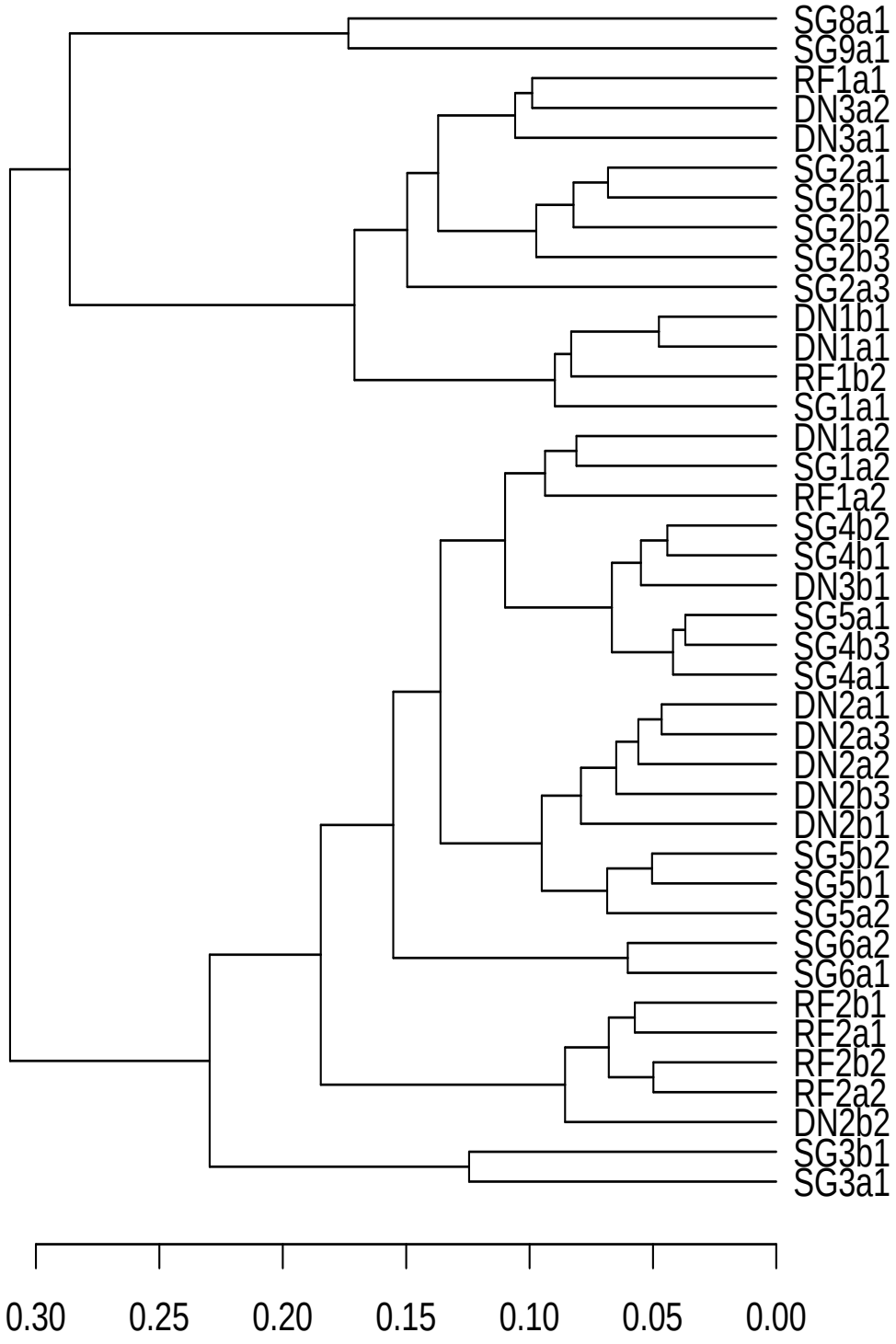


Figure 3.43. UPGMA tree of the bacterial phyla populations of 40 sediment samples from the SG-DN river system, using the Bray Curtis similarity index. The most 18 abundant phyla were selected for the analysis.

3.5.1.3.2. At the genus level:

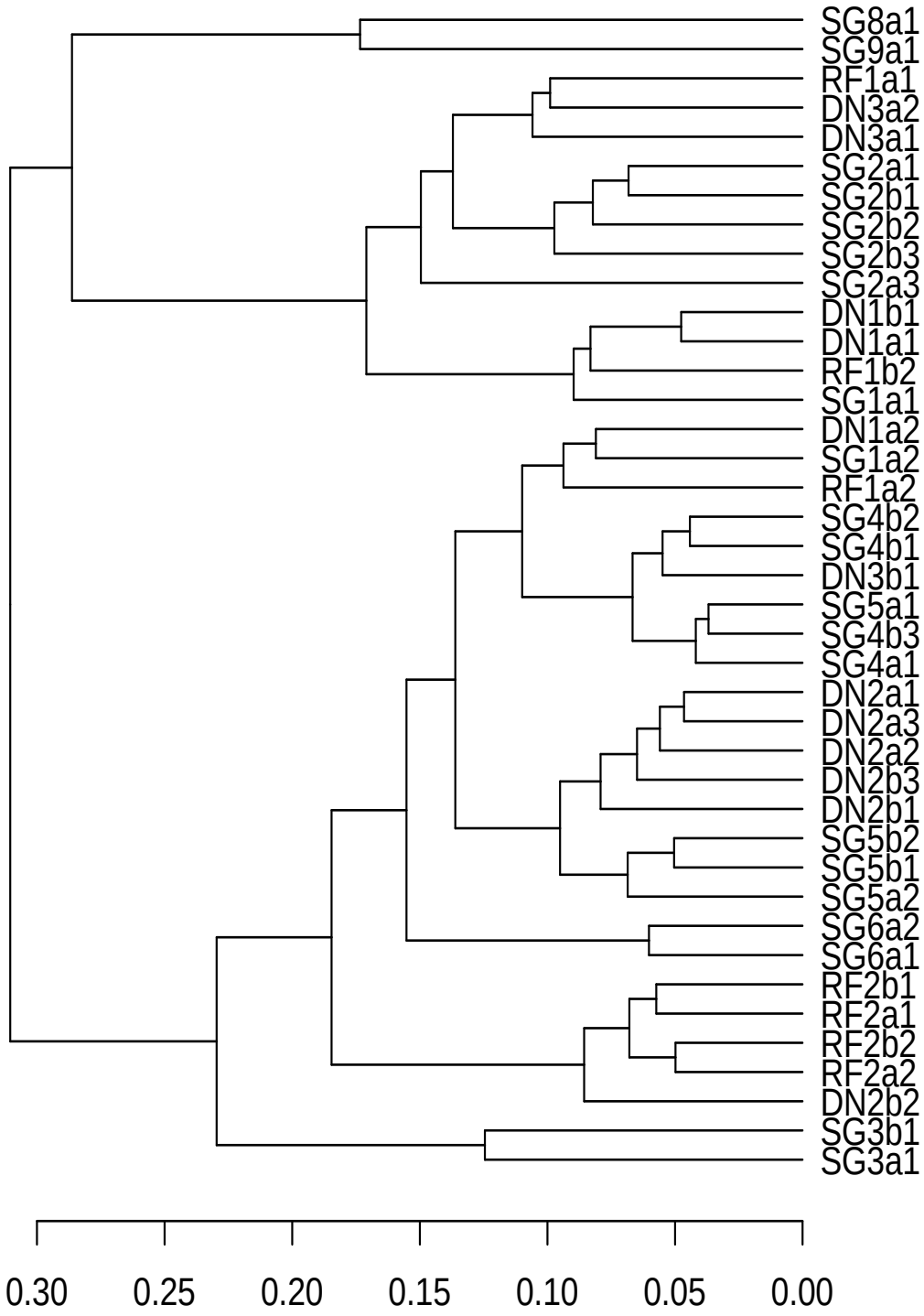


Figure 3.44. UPGMA tree of the bacterial genera populations of 40 sediment samples from the SG-DN river system, using the Bray Curtis similarity index. The most 13 genera abundant were selected for the analysis.

A UPGMA tree, based on the phyla and genera abundance of the 40 samples, was constructed using the Bray-Curtis index to better understand the population relationships among the samples (**Fig. 3.44** & **3.45**). The results showed that UPGMA trees for phyla and for genera trees are very similar. The UPGMA trees separated the 40 samples into two large groups.

The 1st group includes samples RF1 (a1, b2), SG1a1, SG2 (a1, a3, b1, b2, b3) which belong to the SaiGon river; the samples DN1 (a1, b1), DN3 (a1, a2) which belong to the DongNai river and the samples, SG8a1 & SG9a1, which belong to the canals.

The 2nd group includes samples RF2 (a1, a2, b1, b2), DN1a2, DN2 (a1, a2, a3, b1, b2, b3), DN3b1, SG4 (a1, b1, b2, b3), which belong to the DongNai river; the samples RF1a2, SG1a2, SG6 (a1, a2), SG3 (a1, b1) which belong to the SaiGon river, and the samples SG5 (a1, a2, b1, b2) which are from the intersection location.

In summary, the both UPGMA trees showed that the samples SG8a1 & SG9a1 formed and outgroup among samples, as did SG3a1 & SG3b1. The differences in the phyla communities do not seem to be influenced by the right versus the left side of the river, except for the samples of locations RF2 and SG2.

3.5.2. PAHs & *Fecal coliforms* correlation versus bacterial population:

3.5.2.1. Pearson correlation of PAHs and Fecal coliforms versus bacterial population:

3.5.2.1.1. At the phyla level:

Table 3.36: Pearson correlation of the most 17 abundant phyla versus chemical (PAHs) analytes & biological (*Fecal coli&E.coli*) analytes for the first sample of left side (a1) and the first sample of right side (b1) of 13 locations. There are 21 samples, sample RF1b1 was eliminated through the sequences normalization process.

	Un Classified	Proteo bacteria	Chloro flexi	Nitro spirae	Acido bacteria	Bacter oidetes	Actino bacteria	Amini cenantes	Chlorobi	Firmicutes	Gemmati monadetes	Latesci bacteria	Plancto mycetes	Spiro chaeta e	TA06	Cyano bacteria	Verruco microbia	Elusi microbia
Naphthalene	-0.142	-0.129	0.224	-0.489	0.039	0.253	0.044	0.002	-0.333	0.363	0.054	0.102	-0.233	0.069	-0.060	0.406	-0.049	0.002
Anthracene	-0.347	-0.134	0.519	-0.434	-0.237	-0.047	-0.220	-0.257	-0.449	0.673	-0.343	-0.336	-0.586	0.129	-0.235	0.387	-0.151	0.078
Fluoranthene	-0.352	-0.105	0.510	-0.397	-0.242	-0.102	-0.225	-0.212	-0.457	0.614	-0.365	-0.378	-0.540	0.171	-0.214	0.290	-0.116	0.109
Pyrene	-0.383	-0.086	0.504	-0.416	-0.256	-0.107	-0.226	-0.225	-0.459	0.634	-0.370	-0.382	-0.559	0.148	-0.239	0.333	-0.146	0.090
Perylene	**0.703**	**-0.620**	0.478	0.238	-0.371	-0.295	-0.586	0.549	-0.262	0.070	-0.480	-0.239	0.411	0.682	0.657	-0.242	-0.158	0.471
Benzo[a]anthracene+Chrysene	-0.345	-0.158	0.553	-0.408	-0.290	-0.110	-0.235	-0.203	-0.455	0.651	-0.382	-0.358	-0.548	0.154	-0.226	0.380	-0.163	0.116
Benzo[b&k]fluoranthene	-0.383	-0.117	0.527	-0.426	-0.259	-0.088	-0.214	-0.221	-0.448	0.649	-0.382	-0.376	-0.571	0.139	-0.239	0.370	-0.139	0.114
Total PAHs	-0.180	-0.279	0.640	-0.361	-0.336	-0.152	-0.360	-0.077	-0.521	0.650	-0.466	-0.403	-0.447	0.327	-0.060	0.278	-0.179	0.216
Fecal coli	-0.052	0.161	-0.268	-0.012	0.210	0.389	0.121	-0.191	0.125	0.158	0.183	0.023	-0.015	-0.095	-0.004	-0.091	0.131	-0.191
E.coli	-0.162	0.136	-0.250	0.338	0.055	0.087	0.210	-0.239	0.334	-0.041	0.126	0.000	0.052	-0.195	-0.197	-0.112	0.137	-0.277

Note: *Synergistetes* was eliminated because the proportion because its relative abundance, which is >1%, was only presented in 2 out of 40 samples.

3.5.2.1.2. At the genus level:

Table 3.37: Pearson correlation of the most 13 abundant genera versus chemical (PAHs) analytes & biological (*Fecal coli&E.coli*) analytes for the first sample of left side (a1) and the first sample of right side (b1) of 13 locations. There are 21 samples, sample RF1b1 was eliminated through the sequences normalization process.

	Uncultured Anaerolineaceae	Uncultured Nitrosomonadaceae	Uncultured Nitrospiraceae	Nitrospira	Spirochaeta	Anaeromyxobacter	Dechloromonas	Uncultured Alcaligenaceae	Uncultured Comamonadaceae	Uncultured Rhodocyclaceae	Acidiferrobacter	Uncultured Hydrogenophilaceae	Uncultured Syntrophaceae
Naphthalene	0.393	-0.312	-0.464	-0.251	-0.125	-0.188	0.010	-0.179	-0.168	-0.071	-0.194	-0.038	0.481
Anthracene	0.702	-0.383	-0.302	-0.252	-0.324	-0.198	0.446	-0.151	-0.143	0.275	-0.194	-0.011	0.148
Fluoranthene	0.635	-0.321	-0.262	-0.240	-0.292	-0.132	0.470	-0.108	-0.128	0.290	-0.204	-0.079	0.093
Pyrene	0.658	-0.323	-0.302	-0.235	-0.315	-0.117	0.486	-0.082	-0.128	0.305	-0.201	-0.047	0.127
Perylene	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058	-0.058
Benzo[a]anthracene+Chrysene	0.714	-0.386	-0.299	-0.225	-0.307	-0.201	0.440	-0.161	-0.137	0.253	-0.189	-0.012	0.167
Benzo[b&k]fluoranthene	0.689	-0.360	-0.327	-0.229	-0.324	-0.162	0.478	-0.129	-0.133	0.303	-0.209	-0.024	0.150
Total PAHs	0.666	-0.471	-0.105	-0.295	-0.120	-0.239	0.369	-0.202	-0.261	0.188	-0.240	-0.164	0.204
Fecal coli	-0.158	-0.058	-0.146	0.080	-0.049	0.008	0.025	-0.008	0.164	0.204	0.070	-0.042	0.006
E.coli	-0.100	0.113	-0.088	0.428	-0.160	-0.058	-0.109	-0.101	0.504	-0.073	0.372	0.240	-0.133

Note: *Leptolinea* was eliminated because its relative abundance, which is >1%, was only presented in 3 out of 40 samples.

Longilinea was eliminated because its relative abundance, which is >1%, was only presented in 2 out of 40 samples.

Acidobacteriaceae (Subgroup 1) was eliminated because its relative abundance, which is >1%, was only presented in 6 out of 40 samples.

3.5.2.1.3. At OTU & Shannon level:

Table 3.38: Pearson correlation of the OTUs and Shannon index versus chemical (PAHs) analytes & biological (*Fecal coliform*) analytes for the first sample of left side (a1) and the first sample of right side (b1) of 13 locations. There are 21 samples, sample RF1b1 was eliminated through the sequences normalization process.

	OTUs	Shannon
Naphthalene	-0.022	-0.064
Anthracene	-0.505	-0.465
Fluoranthene	-0.534	-0.467
Pyrene	-0.550	-0.492
Perylene	-0.038	0.050
Benzo[a]anthracene+Chrysene	-0.536	-0.493
Benzo[b&k]fluoranthene	-0.538	-0.489
Total PAHs	-0.528	-0.453
Fecal coli	0.126	0.106

3.6. Mean &Standard Deviation of 40 samples*:*

3.6.1 At the phyla level:

Table 3.39: Mean & Standard Deviation of the most 17 abundant phyla for total 40 sediment samples.

	No Relative	Proteobacteria	Chloroflexi	Nitrospirae	Acidobacteria	Bacteroidetes	Actinobacteria	Aminicenantes	Chlorobi	Firmicutes	Gemmatimonadetes	Latescibacteria	Planctomycetes	Spirochaetae	TA06	Cyanobacteria	Verrucomicrobia	Elusimicrobia
Mean	21.3	29.2	18.1	7.5	4.8	2.6	1.5	1.3	1.8	0.8	0.7	0.6	2.4	1.6	0.8	1.0	0.7	0.5
Standard Deviation	4.0	10.3	8.5	4.2	2.2	2.0	0.8	1.4	0.8	0.6	0.4	0.4	0.7	0.8	0.6	2.6	0.9	0.4

3.6.2. At the genus level:

Table 3.40: Mean & Standard Deviation of the most 13 abundant genera for total 40 sediment samples.

	Uncultured Anaeroline aceae	Uncultured Nitrosomonadaceae	Uncultured Nitrospira ceae	Nitrospira	Spirochaeta	Anaeromyxobacter	Dechloromonas	Uncultured Alcaligenaceae	Uncultured Comamonadaceae	Uncultured Rhodocyclaceae	Acidiferrobacter	Uncultured Hydrogenophilaceae	Uncultured Syntrophaceae
Mean	7.9	2.0	2.9	3.2	1.3	0.9	1.0	1.1	1.0	0.9	0.6	0.4	0.5
Standard Deviation	2.4	1.4	2.3	3.6	0.8	1.2	1.3	1.6	0.8	0.8	0.8	0.4	0.4

3.6.3. At OTU & Shannon level:

Table 3.41: Mean &Standard Deviation of OTUs & Shannon index

	OTUs	Shannon
Mean	2078.3	7.3
Standard Deviation	209.9	0.3

CHAPTER 4: DISCUSSION

4.1. Chemical analysis of the SG-DN river system (February 2012):

4.1.1. Total Organic Carbon (TOC):

TOC levels in surface sediment samples taken in May 2004 (7 locations SGR1-SGR8) ranged from 1.9% to 3.8% with an average of 2.6% (**269**). Compared to our data, TOC levels increase in the SaiGon river from 2004 to 2012 in the factor of 2. The comparison showed that there is an accumulation of organic pollutants in the SaiGon river. The increasing TOC levels are probably due to the increasing of population, urbanization and the industrial activities in HCMC, Dong Nai and Binh Duong provinces. In May 2007, the average TOC levels were highest in the canals (4.0%), decreased in the river (2.4%) and lowest in the estuary (1.2%).

TOC level varies among other rivers in the world. TOC levels in the surface sediment river of urban estuary (location Ballona Creek, South Californa, USA., samples were take in October 2008) was from 0.31-2.95% (**270**). In Lower Mekong River Basin (samples taken on December 2005), the average of TOC level in the surface sediment is 0.41- 0.55% (**31**). The average TOC level from the canals in Vientiane, Laos was 3.5%.

Note :

a: Some of the TOC avarage that mentioned above were calculated by the author (me).

4.1.2. Heavy Metals:

Contamination of heavy metals in the SG-DN river system showed that no samples exceeded the PEL criteria. However, the concentration of heavy metals in several samples exceeded the TEL criteria. Five samples have Zn concentrations that are above the TEL. Similarly, Cu concentration in 6 samples exceeded the TEL. The concentration of Pb and Cr in 5 samples reached the TEL. Ni concentrations in most of the samples are below the Guideline criteria. The contamination of the heavy metals should be further monitored in the SG-DN river system

4.1.2.1. Compared with other rivers:

Comparison is included in order to see how the concentration of several heavy metals in sediment vary in different regions in the world, from the industrial basin of China (North Asia), Morocco (Africa) and Thailand (Southern Asia).

- Pb concentrations in the SG-DN rivers were 17.2-43.4 $mg.kg^{-1}$. Pb levels vary among rivers in the world. The average Pb concentration of a coastal industrial basin polluted river in China ranged from 112.28-2431.09 $mg.kg^{-1}$ (**271**), of Bas

Oum Erbia (Morocco, samples taken in 2002) is 0.4 mg.kg^{-1}, of Day River (Morocco, samples taken in 2010) is 140.35 mg.kg^{-1} **(272**).

- Ni concentrations in the SG-DN rivers were 22.4-83.2 mg.kg^{-1}. Ni levels also vary rivers in the world. The average Ni concentration of the coastal industrial basin polluted river in China ranged from 31.49-812.91 mg.kg^{-1} (**271**). The Ni concentration of eastern coast of the Gulf of Thailand is 79.9 mg.kg^{-1} and of Laem Chabang (Chonburi Province, Thailand) is 0.64 mg.kg^{-1} (**273**).
- Cu concentrations in the SG-DN rivers were 13.9-57.9 mg.kg^{-1}. The average Cu concentration of the coastal industrial basin polluted river in China ranged from 50.9-1533.33 mg.kg^{-1} (**271**). In Morocco, the concentration of Cu ranged from 2 mg.kg^{-1} (from Moulay Bousselham Lagoon, 2010) to 723 mg.kg^{-1} (Martil River from northeast of Tetouan city, 2003) (**272**). In Southern Asia, the concentration of Cu ranged from 14.4 mg.kg^{-1} (Prasae River, Chanthaburi province, Thailand) to 103 mg.kg^{-1} (Eastern Coast of the Gulf of Thailand) (**273**).
- Cr concentrations in the SG-DN rivers were 21.1-54.5 mg.kg^{-1}. Cr concentrations were 7.0 mg.kg^{-1} (Mghogha river, north of Morocco, 2009) and 250.4 mg.kg^{-1} (Hindon river, Saharanpur district from upper Shivalik to lower Himalayan range, India, 2009) (**272**).
- Zn concentrations of the SG-DN rivers were 65.0-168.0 mg.kg^{-1}. The average Cu concentration of the coastal industrial basin polluted river in China ranged from 256.78-6546.57 mg.kg^{-1} (**271**). In Morroco, Zn concentration were 4 mg.kg^{-1} and 1190 mg.kg^{-1}, respectively (Nador Lagoon from Nador city, 2009) (**272**). In Thailand, Zn concentration was from 7.48 mg.kg^{-1} (Laem Chabang and Coast of the Gulf of Thailand) to 131 mg.kg^{-1} (Bangpakong river from Prachinburi Province, 2004) (**273**).

4.1.2.2. Compared with previous study in the SG-DN rivers (235) :

Heavy metals, including Cu, Pb, Zn, Cr and Ni, were surveyed before in the SG-DN rivers system in 1997-1998 (sediments were taken every three months between November 1997 and December 1998, always at low tide). A total of 10 sampling sites were chosen, with 5 sites in the SaiGon river and 5 sites in the DongNai river. The concentration of Cu ranged from 19.38-48.41 mg.kg^{-1} with an average concentration of 32.04 mg.kg^{-1}. The Cu concentrations of the SG-DN rivers in 2012 were 13.9-57.9 mg.kg^{-1}, with an average concentration of 45.3 mg.kg^{-1} increased compared to those in 1998.

The concentrations of Pb ranged from 13.15-36.68 $mg.kg^{-1}$ with an average concentration of 35 $mg.kg^{-1}$. The Pb concentrations of the SG-DN rivers in 2012 were from 17.2-43.4 $mg.kg^{-1}$, with an average 29.3 $mg.kg^{-1}$, slightly decreased compared to those in 1998.

The concentration of Zn ranged from 43.38-201.74 $mg.kg^{-1}$ with an average concentration of 104.52 $mg.kg^{-1}$. The Zn concentrations of the SG-DN rivers in 2012 were from 65.0-168.0 $mg.kg^{-1}$, with an average concentration of 124.4 0 $mg.kg^{-1}$, slightly increased compared to those in 1998.

The concentration of Cr ranged from 21.61-118.3 $mg.kg^{-1}$ with an average concentration of 48.77 $mg.kg^{-1}$. The Cr concentrations of the SG-DN rivers in 2012 were from 21.1-54.5 $mg.kg^{-1}$, with an average concentration of 42.9 $mg.kg^{-1}$, slightly decreased compared to those in 1998.

The concentration of Ni ranged from 22.86-115.45 $mg.kg^{-1}$ with an average concentration of 55.71 $mg.kg^{-1}$. The Ni concentrations of the SG-DN rivers in 2012 were from 22.4-83.2 $mg.kg^{-1}$, with an average of 52.8 $mg.kg^{-1}$, slightly decreased compared to that in 1998.

Thus, during 14 years, the heavy metal concentrations in the SG-DN river sediment do not change to a significant degree.

4.1.3. PAHs:

The concentration of total PAHs is highest in the SG4 sample. SG4 sediment samples were taken near the Cat Lai Industrial Park, and the Fuel – Oil Factory. The total PAHs in SG4 probably indicate the activity of the oil production nearby (Fig. 3.2.3&3.2.4).

The concentration of total PAHs is also high in SG3 sample. SG3 sediment was taken near the two large industrial parks, which are Linh Trung and Binh Chieu, and surrounding by other six industrial parks (IP), including Viet Huong IP, VN-Singapore IP, Song Than II IP, Dong An IP, Binh Duong IP, Song Than I IP.

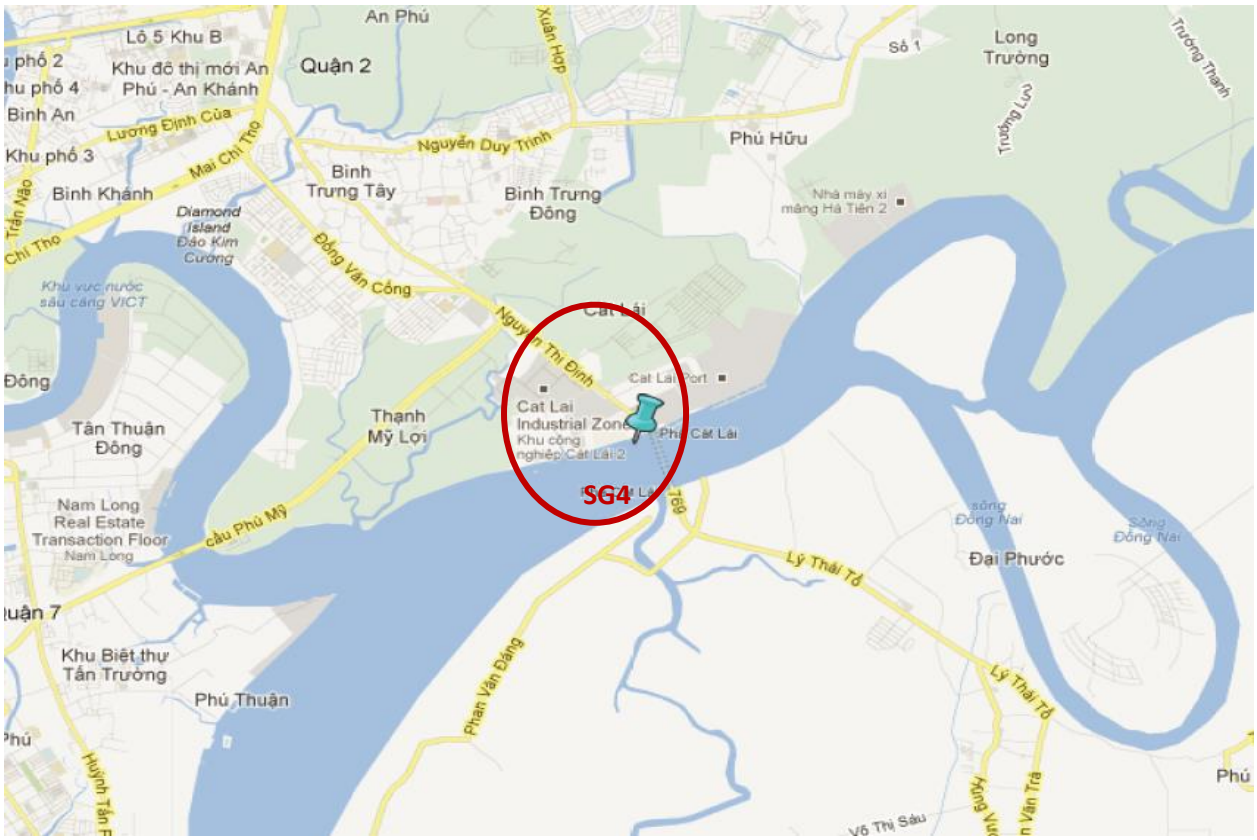


Figure 4.1. The surrounding area of SG4 location (map was taken on Google December 2012).

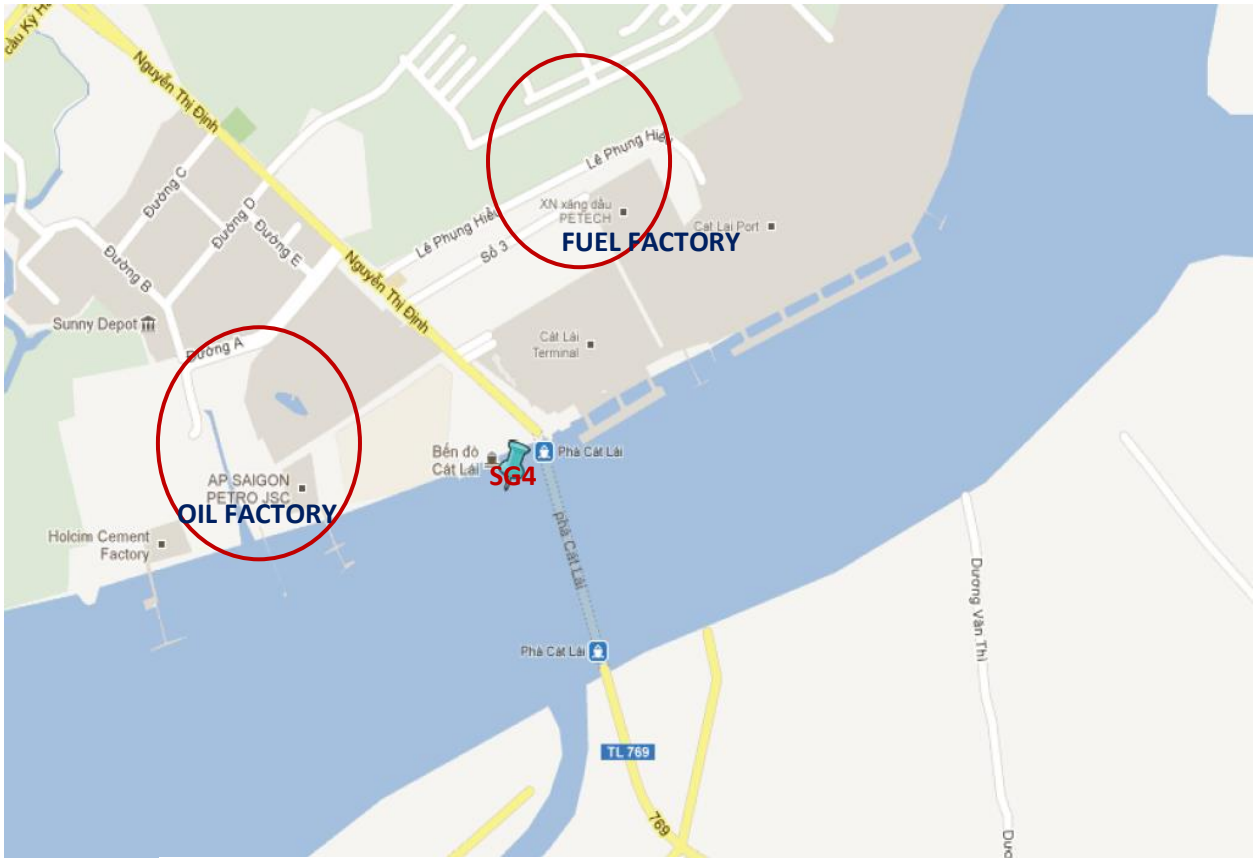


Figure 4.2. Oil & Fuel Factory nearby location SG4.

U.S Environmental Protection Agency (EPA) includes the following PAH compounds that could cause cancer.

1. Benz [a] anthracen,
2. benzo [a] pyrene,
3. benzo [b] fluoranthene,
4. benzo [k] fluoranthene,
5. chrysene,
6. dibenz [a, h] anthracen
7. indeno [1,2,3-cd] pyrene.

4.1.3.1.Compared with other rivers and soils in the world:

Naphthalene:

The naphthalene concentrations of the SG-DN rivers are from 25.0-133.0 $ng.g^{-1}$, higher than those of the Soltan Abad River (Iran) which are from 7.9-18.8 $ng.g^{-1}$ (**274**); and much higher than those (calculated NBC) of urban soils from Greater London (UK), with 0.34 $mg.kg^{-1}$ for urban, 0.23 for semi-urban, 0.29 $mg.kg^{-1}$ for urban + semi-urban (**275**). For the Soil quality guideline for environmental health SQG_E based on the lowest of the available environmental health guidelines (with soil contact, soil and food ingestion, or protection of freshwater life), the naphthalene concentrations that are allowed for Agricultural and Residential use are 8.8 and 8.8 $mg{\cdot}kg^{-1}$, respectively. There is no napthalene value for Parkland, Commercial and Industrial land use (**276**, page 161). The results showed that the naphthalene concentrations of the SG-DN rivers are much higher than other areas in the world, indicating high pollution levels of this compound in the river.

Anthracene:

The anthracene concentrations of the SG-DN rivers are 36-65 $ng.g^{-1}$, higher than those of the Soltan Abad River (Iran) which are 7.3-15.3 $ng.g^{-1}$ (**274**); and much higher than the those (calculated NBC) of ubran soils from Greater London (UK), with 1 $mg.kg^{-1}$ for urban, 0.6 $mg.kg^{-1}$ for semi-urban, 0.81 $mg.kg^{-1}$ for urban + semi-urban (**275**). For the Soil quality guideline for environmental health SQG_E based on the lowest of the available environmental health guidelines (with soil contact, soil and food ingestion, or protection of freshwater life), the anthracene concentrations that are allowed for Agricultural,

Residential or Parkland, Commercial and Industrial land use are 2.5, 2.5, 32 and 32 $mg \cdot kg^{1}$, respectively (**276**, page 161). The results showed that the anthracene concentrations of the SG-DN rivers are much higher than other areas in the world, indicating the high pollution level of this compound in the river.

Fluoranthene:

The fluoranthene concentrations of the SG-DN rivers are 28-54 $ng.g^{-1}$, quite lower than those of Soltan Abad River (Iran) which are 17.5-92.6 $ng.g^{-1}$ (**274**). The anthracene concentration (calculated NBC) of urban soils from Greater London (UK), which are 12 $mg \cdot kg^{-1}$ for urban, 5.3 for semi-urban and 9.7 $mg \cdot kg^{-1}$ for urban + semi-urban (**275**). For the SQG_E based on the lowest of the available environmental health guidelines (with soil contact, soil and food ingestion, or protection of freshwater life), the fluoranthene concentrations that are allowed for Agricultural, Residential or Parkland, Commercial and Industrial land use are 50, 50, 180 and 180 $mg \cdot kg^{-1}$, respectively (**276**, page 161). The results showed that the fluoranthene concentrations of the SG-DN rivers are higher than the background soil, indicating the human activities involved (**276**, page 161); but equal to or lower compared with Agricultural, Residential or Parkland, Commercial and Industrial land, indicating the fluoranthene concentrations of the SG-DN river may be acceptable.

Pyrene:

The pyrene concentrations of the SG-DN rivers are 29-69 $ng.g^{-1}$, lower than those of the Soltan Abad River (Iran) which are 16.2- 55.4 $ng.g^{-1}$ (**274**); and higher than the pyrene concentration (calculated NBC) of urban soils from Greater London (UK), which are 11 $mg \cdot kg^{-1}$ for urban, 5 $mg \cdot kg^{-1}$ for semi-urban and 8.4 $mg \cdot kg^{-1}$ for urban + semi-urban (**275**). For Superseded Interim Soil Quality Criteria (CCME 1991), the pyrene concentrations that are allowed for Agricultural, Residential or Parkland, Commercial and Industrial land use are 0.1, 1, 10 and 10 $mg \cdot kg^{-1}$, respectively (**276**, page 161). The results showed that pyrene concentrations of the SG-DN rivers are higher than the background soil and the Superseded Interim Soil Quality Criteria, indicating the human activities involved.

Benzo[a]pyrene:

The benzo[*a*]pyrene concentrations of the SG-DN rivers are 20-43 $ng.g^{-1}$, similar to those of the Soltan Abad River (Iran) which are 11.5 - 36.7 $ng.g^{-1}$ (**274**); higher than the those (calculated NBC) of urban soils from Greater London (UK), which are 8.7 $mg \cdot kg^{-1}$ for urban, 4.6 for semi-urban and 7.0 $mg \cdot kg^{-1}$ for urban + semi-urban (**275**). For the SQG_E based on the lowest of the available environmental health guidelines (soil contact, soil and

food ingestion, or protection of freshwater life), the anthracene concentration that is allowed for Agricultural, Residential or Parkland, Commercial and Industrial land use are 0.6 and 0.6 mg·kg^{-1}, respectively (**276**, page 161). The results show that benzo[*a*]pyrene concentrations of the SG-DN rivers are much higher than the background soil and the SQG_E levels, indicating the human activities involved and the pollution level of benzo[*a*]pyrene in the SG-DN rivers.

Dibenz[a,h]anthracene:

The dibenz[*a,h*]anthracene concentrations of the SG-DN rivers are 20-43 ng.g^{-1}, similar to those of the Soltan Abad River (Iran) which are 11.5 - 36.7 ng.g^{-1} (**274**). The dibenz[*a,h*]anthracene concentrations (calculated NBC) of urban soils from Greater London (UK) are 6.9 mg·kg^{-1} for urban, 4.4 for semi-urban and 6.0 mg·kg^{-1} for urban + semi-urban (**275**). For the SQG_E based on the lowest of the available environmental health guidelines (soil contact, soil and food ingestion, or protection of freshwater life), the dibenz[*a,h*]anthracene concentration that is allowed for Agricultural, Residential or Parkland, Commercial and Industrial land use are 0.6 and 0.6 mg·kg-1, respectively (**276**, page 161). The results show that dibenz[*a,h*]anthracene concentrations of the SG-DN rivers are much higher than the background soil and the SQG_E levels, indicating the human activities involved and pollution levels of benzo[*a*]pyrene in the SG-DN rivers.

Benzo[g,h,i]perylene:

The benzo[g,h,i]perylene concentrations of the SG-DN rivers are from 19 mg.kg^{-1} to 65 mg.kg^{-1}, higher than those in background soils that came from different types and regions all over the world (0.0005 to 0.67 mg.kg^{-1}), and exceed the Netherlands 'Maximum Permissible Concentration' for sediment (7.5 mg.kg^{-1}) (**276**) and groundwater protection soil threshold values based on drinking water-related cancer risks (6.8 mg·kg^{-1}) (**276**). Moreover, it is similar to the concentrations in 5cm-surface sediments of the Soltan Abad River (Iran) during April 2013 to March 2014, which are from 20.7 to 30.22 ng.g^{-1} (**274**).

Indeno[1,2,3-cd]pyrene:

The indeno[1,2,3-cd]pyrene concentrations in the SG-DN rivers are from 23 to 57 ng g^{-1} and higher than those in Soltan Abad River (Iran) (3.3-12.9 ng.g^{-1}) and those of urban soils from Greater London (UK), which are 6.8 mg.kg^{-1} for urban and 5.2 mg.kg^{-1} for urban + semi-urban.

Overall:

The Soltan Abad River (Iran) is located in the south of Shiraz, passing through the industrial town of Shiraz with numerous factories and industries (industrial materials and chemical products, rubber and plastics, metal artifacts, etc.) which all have organic matter like PAHs. They also have urban sewage and agricultural lands that serve as other PAHs sources (**274**). The SG-DN rivers share same characteristics with the Soltan Abad River about numerous industrial parks and factories located across the river. The area of the SG-DN rivers in this study locate in the highest population area of the country (HCMC, Dong Nai and Binh Duong provinces), with urban sewage and also some small-scale agricultural lands. The PAHs concentrations in sediment of the SG-DN rivers are similar (slightly lower or higher) in those of the Soltan Abad River. The PAHs concentrations of the SG-DN rivers are much higher than the background soil, probably come from anthropogenic sources of the area.

4.1.4. PCBs:

Previous study of PCBs in May 2004 (**269**) showing that PCBs were highest in canals, with an average concentration of 8.1 $ng.g^{-1}$, the decreased towards the river with an average concentration of 6.8 $ng.g^{-1}$ and lowest in the estuary with an average concentration of 0.9 $ng.g^{-1}$. The PCBs concentration in the SaiGon river (location SGR1-SGR8) ranged from 1.8-8.8 $ng.g^{-1}$ with an average concentration of 6.3 $ng.g^{-1}$.

4.2. Microbial analysis of the sediments from the SaiGon-DongNai river system (August 2012):

Project: Analyzing the effect of industrial and urban polluted zones on microbial diversity in the Sai Gon-Dong Nai river sediment (Vietnam).

Though the SG-DN river system, a moderate polluted river system, has been well studied with various chemical contaminants, its overall resident bacterial populations have remained largely unknown. Using pyrosequencing of the V3-V1 region of 16S rDNA, we were able to access the sedimental bacterial communities of the SG-DN river system for the first time, regarding the geological ecosystems of the river. Understanding the bacterial components of polluted aquatic environments, such as the SG-DN river system, will increase our knowledge about the overall structure of bacteria and function of their ecological roles in polluted environments, therefore enhancing our ability to monitor the quality of public health which depend on the quality of the river.

River sediments contain a large variety of environmental contaminants and play a key role in the ecological status of aquatic ecosystems. Contaminants that were absorbed by sediments and supended solids may contribute directly or, after remobilization, to an adverse ecological and chemical status of surface water and change the diversity of natural bacterial components of the sediment **(277)**.

4.2.1. At the phyla level:

The bacterial communities of the SG-DN river sediments are dominated with the phyla *Proteobacteria, Chloroflexi, Nitrospirae* and *Acidobacteria*. Phyla *Bacteroidetes, Actinobacteria, Aminicenantes, Chlorobi*, *Planctomycetes, Verrucomicrobia, Spirochaetae* and *Firmicutes* appeared with less abundance among the samples of the river. Similarly, a study of bacterial compositions in an urban river impacted by different pollutant sources by MarkIbekwe et al. (2016), showed that phyla *Proteobacteria, Bacteroidetes, Acidobacteria,* and *Actinobacteria* were dominated in all sediment samples (**278**). Another study in urban park soils of 16 representative Chinese cities using the pyrosequencing revealed the 6 dominant phyla present in all samples were *Proteobacteria, Actinobacteria, Acidobacteria, Planctomycetes, Chloroflexi* and *Bacteroidetes* (**279**). Members of β-*Proteobacteria*, *ε-proteobacteria*, *Acidobacteria*, *Bacteroidetes* and *Verrucomicrobia* were also found in bacterial composition of an urban river in the North West Province, South Africa (**280**). The bacterial composition of the SG-DN river sediments shared similar characteristics with other urban sediments from different regions (**278, 279, 280**, **281**).

The bacterial community of the SG-DN river sediments are most dominated by the phylum *Proteobacteria,* ranging from 10.88% to 61.63% before the normalization and 8.4%-56.1% after the sequencing number normalization. Ligi et al. (2013) found that *Proteobacteria* composed 22.7%-59.2% of all the sequences in freshwater sediments (**282**). Roesch et al. (2007) reported that more than 40% of the soil sequences were *Proteobacteria* (**283**). In a study of sewage treatment plants, *Proteobacteria* was the most abundant phylum in all of the sludge samples, accounting for 35% to 65% of the community abundance (**138**).

Phylum *Chloroflexi* ranked the second most abundant with the proportion from 8.58%-46.43% and from 7.4%-37.9% before and after number sequencing normalization, respectively. The phylum *Chloroflexi* is a large phylum containing members of bacteria associated with various metabolic features, one of which is the anoxygenic photosynthetic activity (**285**). A study of the bacterial diversity in urban lakes sediments by T-RFLP showed that *Chloroflexi* were the most dominant bacterial group in the clone library with

proportion of 21.7 % of the clones, which was partly associated with its higher total nitrogen and organic matters concentrations (**286**). *ε-Proteobacteria* and *Chloroflexi* abundance, which comprised 44.9% of total clones, distinguished between polluted and unpolluted sediment samples of the flora bacterial communities (**287**)**.** Filamentous *Chloroflexi*, the green non-sulfur bacteria, were also found to be abundant in wastewater treatment processes with biological nutrient removal (**288**). Members of the bacterial phylum *Chloroflexi* are common and highly diverse in sediment. Genomic analyses provide new evolutionary boundaries for obligate organohalide respiration of these members. The potential roles of *Chloroflexi* in sediment carbon cycling beyond organohalide respiration were shown, including respiration of sugars, fermentation, CO_2 fixation, and acetogenesis with ATP formation by substrate-level phosphorylation (**289**).

The third most dominant phylum was *Nitrospirae* ranging, from 0.56%-18.23 % before the normalization and 0.4%-16.8% after the sequences number normalization. Unlike *Proteobacteria*, phylum *Nitrospirae* are rarely found wiith abundant proportions in various sediment niches (**138, 278, 279 280, 281, 282, 283**). The *Nitrospirae* play a key role in the nitrogen cycle, in which its genus *Nitrospira* is a nitrifier (**284**). Study of profiling bacterial communities associated with sediment-based aquaculture bioremediation systems under contrasting redox regimes showed that the phylum *Nitrospirae*, the candidate divisions AncK6, GAL15, SBR1093, TM7 and the *Proteobacteria* sub-class TA18 were only present in oxic sediments (**290**). Tag pyrosequencing of bacterial 16S rRNA genes revealed significant effects of effluent on sedimental bacterial compositions, with an increase in abundance of *Nitrospirae* and *Sphingobacteriales* sequences in the Chicago metropolitan region (**291**).

Actinobacteria are widely distributed in both terrestrial and aquatic (including marine) ecosystems and especially in soil, where they play a crucial role in the recycling of refractory biomaterials by decomposition and humus formation. Furthermore, *Actinobacteria* members have adopted different lifestyles, and can be pathogens such as *Corynebacterium*, *Mycobacterium*, *Nocardia*, *Tropheryma*, and *Propionibacterium*, soil inhabitants including *Streptomyces*, plant commensals such as *Leifsonia*, or gastrointestinal commensals with *Bifidobacterium* (**292**).

4.2.2. At the genus level:

Ecological meaning of genera in the SG-DN river system:

The uncultured genera of several families had high proportions among the samples, including *Uncultured Anaerolineaceae*, *Uncultured Nitrosomonadaceae*,

Uncultured Nitrospiraceae. However, there are few studies of the uncultured genera. A study of bacterial community in heavy metal polluted soils, using 16S rDNA pyrosequencing analysis by Marcin et al. (2014), mentioned that the four most abundant genera were uncultured members of *Acidobacteriaceae*, *Gemmatimonadaceae*, *Nitrosomonadaceae*, and *Xanthobacteraceae* (**293**).

- ***Nitrospira****:*

Members of *Nitrospira* play an important role in nitrification process of the biogeochemical nitrogen cycle. First step in the metabolic process, ammonia oxidizer *Nitrosomonas* oxidizes ammonia into nitrite in aerobic condition, and nitrite is oxidized into nitrate by *Nitrospira* (**294, 295, 296**).

High populations of *Nitrospira* are probably caused by high concentrations of nitrite in sediment samples. Nitrites are often used as corrosion inhibitors in industrial process water and cooling towers (**297**, **298**, **299**). *Nitrospira* was found as a dominant genus at downstream of wastewater treatment plants (WWTP) in the Seine River (**300**). In addition, nitrospira-like nitrite-oxidizing bacteria (NOB) are present in the polluted sediments of Niida River (Hachinohe, Japan) along with ammonia-oxidizing bacteria (AOB) (**301**). Studies of the Seine river showed that there was nitrite accumulation at the downstream stations (Poissy and Posses) during 6 years from 2007 to 2013 compared with other polluted nitrogen: ammonia and nitrate (**302**). The total nitrogen (ammonia, nitrite and nitrate) concentrations of the SaiGon river were lower at the upstream Thu Dau Mot location (1.5 - 1.8 $mg.l^{-1}$), and higher at the downstream Nha Rong Harbor location (2.4 - 3.2 mg^{-1}) (**303**). The results here may lead ti study that can use the *Nitrospira* genus abundance for indicating the pollution levels caused by industrial activities.

- ***Acidiferrobacter****:*

According to List of Prokaryotic names with Standing in Nomenclature (LPSN), genus *Acidiferrobacter* belongs to family *Acidiferrobacteraceae* with 2 other genera that are *Sulfuricaulis* and *Sulfurifustis* (**304**). It is acidophilic and possesses diazotrophs characteristic, which fixes nitrogen gas into a more usable form such as ammonia. It is also able to tolerate of elevated concentrations of many metals typically found in mine-impacted environments. The former name of *Acidiferrobacter thiooxydans* is *Thiobacillus ferrooxidans* (Harrison, A P, Jr, 1982), which was also isolated from forty-year-old coal refuse, the Bevier coal seam, Calloway country (Missouri, USA) (**306**). Other members of the family *Acidiferrobacteraceae*, including *Sulfuricaulis* and *Sulfurifustis*, were isolated

from lake sediments (**307**). The higher classification of this genus is Acidiferrobacterales › Gammaproteobacteria › Proteobacteria, which is order to phylum.

Members of the genus *Acidiferrobacter* have been found in other environments such as marine sediments (**308**), limestone aquifer assemblages (**309**), copper ore bioleaching system (**310**), deep-sea sediments (**311**), mine drainage soil (**312**) and mineral-enriched biochars (**313**).

- ***Uncultured Hydrogenophilaceae:***

Uncultured Hydrogenophilaceae belongs to the family *Hydrogenophilaceae*, a family within the order *Hydrogenophilales*, comprises the genera *Thiobacillus*, *Hydrogenophilus*, *Petrobacter*, *Tepidiphilus*, and *Sulfuricella*. Most members of the family are chemolithotrophic or mixotrophic using various inorganic electron donors such as reduced sulfuric compounds or hydrogen. Members of the family are either mesophilic or moderately thermophilic and have been isolated from various environments, e.g, freshwater, aerobic digesters on water treatment sludge, and hot springs. They are also capable of denitrification with the products such as NO_2 and N_2.

- ***Uncultured Acidobacteriaceae (Subgroup 1):***

Uncultured Acidobacteriaceae (Subgroup 1) belong to the family of *Acidobacteriaceae (Subgroup 1)*, order *Acidobacteriales*, class *Acidobacteria*, phylum *Acidobacteria*. There are 11 genera in this family, including the uncultured ones. Sequences present in *Uncultured Acidobacteriaceae (Subgroup 1)* which origin is from the environment such as soil (https://www.arb-silva.de/browser/ssu-122/EU780183/)

4- Uncultured Anaerolineaceae:

Members of family *Anaerolineaceae* were found to be abundant in anaerobic digesters treating waste activated sludge as primary fermenters (**314**).

4.2.3. OTUs, Chao1 and Shannon:

Bacterial community composition of an urban river in the North West Province, South Africa, showed higher richness and evenness at the downstream sites (**319**). Similarly, bacterial community composition of the samples collected upstream from the WWTP discharge were significantly different from that of downstream samples and WWTP effluents (**320**). Comparison of bacterial diversity of polluted and unpolluted sediment by brominated flame retardant revealed that bacterial community structure of polluted sediment was different from the unpolluted sediment sample (**287**).

CHAPTER 5: PERSPECTIVE

5.1. Controlling the toxicity of the SG-DN river system by chemical analyses:

Chemical of the SG-DN river system, including TOC, heavy meals, PAHs and PCBs should be accessed once per year by the national scientists to control the level of toxicity in the river; therefore controlling the waste level from industrial parks which located across the river.

5.2. Factors that affect the bacterial composition of the SG-DN river system:

The bacterial composition reflects the metropolitan characteristics of the SG-DN river system, from upstream to downstream of the river. Methods of analyzing bacterial components using 16S rDNA database, classification methods, 16S rDNA regions that were analyzed and 16S rDNA copy number.

5.2.1. 16S rDNA copy number:

How the 16S rDNA copy numbers vary in each bacterial species and how diverse (the homology level among those 16S DNA sequences) are they?

To know that bacteria species contain how many 16S rDNA copy numbers in their cell, genomic sequence of that organism is required. Annotating 16S rDNA in a genome often depends on two methods: one is sequence similarity searching (called BLAST) and second is program RNAmmer using hidden Markov (HMMs) models based on structural alignments (**321**). With these methods, 16S rDNA gene copy numbers for each prokaryote (bacteria & archaea) were defined. Pei et al. (2010) examined the diversity of 16S rDNA in each genome of 883 prokaryotes representing 568 unique species, of which 425 species contained 2 to 15 copies of 16S rRNA genes per genome. The diversity of the 16S rRNA genes within a genome were calculated by the number of revealed mismatches and insertions divided by the total number of positions, including gaps in the alignment (**322**). An informative list was created with 16S rRNA gene copy numbers for each genome and the diversity level accompanied arcros 25 phyla from *Probacteria* (264 species) to *Acidobacteria* (3 species) (Table S1, **322**). This means that the 16S rDNA copy number information that we have is limited for all the bacteria we can survey in a desired environment.

As described in the “Limit of 16S rDNA based methods” in Introduction, 16S gene copy numbers affect bacterial community abundance analyses (either over- or under-estimate the community). Variation in 16S gene abundances can be caused by both

genomic copy number variation and variation in the abundance of organisms (**323**). Therefore, microbial ecologists who study bacterial community profiles which change according to different physical and chemical environmental characteristics need to take this issue into account.

Kembel et al. (2012) tried to normalize bacterial communities based on 16S gene copy numbers (**321**). A reference database was built by choosing bacterial genomes in which full-length 16S gene sequences are present and genomic 16S gene copy numbers are available. Reference databases were then aligned, masked with PyNAST program and phylogeny constructed. Copy number estimation is calculated by the existence of a phylogenetic signal. The measurement of phylogenetic signal in 16S copy number was performed using the K statistic comparing the amount of signal in a trait to the amount expected under a Brownian motion model of trait evolution. Their method was first linking 16S gene copy numbers, gene abundance, and organismal abundance by developing three sequential algorithms. Second, they estimated organismal relative abundance with the fourth algorithms. To estimate copy number for a novel taxon depending on reference phylogeny with copy number known for all reference taxa, the phylogenetic tree at the common ancestor of the novel taxon was rerooted for its closest relative on the reference phylogeny. 16S copy numbers were predicted at the new root node of the phylogeny. The branch length connecting the root and novel taxon was used to adjust the estimation of predicted copy number.

5.2.2. Does normalization of 16S rDNA useful for bacterial community analysis in urban & industrial polluted sediment samples of the SG-DN river?

To answer this question, we need to see how the bacterial community changes before and after normalization of the sediment samples in different geological, chemical and biological characteristics of the SG-DN river system. Five taxonomic ranks from phyla to genus were observed. Three main ideas, observations were analyzed, one need to see:

1/ Which particular dominant taxa increase or decrease after going through the normalization?

2/ Which particular taxa disappear or appear after going through the normalization?

3/ Does these changes reflect the natural behavior of SG-DN river system?

ANNEX

Annex 1: Characteristics of sampling locations:

A.1. Location SG1:

1-Description:

This is the junction between 2 rivers: the SaiGon river and Thi Tinh river. Because of the flow of the river and there is no way to enter the river bank so that the team just collected only one sample on the Thi Tinh river. This is the border of Ben Cat and Thu Dau Mot. The samples were collected 800 m from the Ong Co Bridge.

The area around the sampling location is used mainly for rice and salad culturing. The river along Thi Tinh- Ong Co bridge to the junction with the SaiGon river is the area of sand transportation, with many transporting ships and equipment present. This is downstream of the Thi Tinh river where industrial parks, industrial plants, densely populated areas, farming areas lying located the river.

2-The coordinate of 2 sides of the location:

- SG1a: N 11^0 02’ 28,46” E 106^0 36’ 18,09”
- SG1b: N 11^0 02’ 24,96” E 106^0 36’ 9,61”

3-Biological replications:

The distance between the sites a1, a2, a3 is very close together so that we extracted total DNA from the a1 and a2 samples.

A.2. Location SG2:

1-Description:

The samples were collected downstream of Ho Phu Water Pumping Station, away from the pumping station 100 m. This is the station of transporting cement of several small boats.

This is one of the water supply resources for Ho Chi Minh City with 300.000 m^3/day We chose this location to investigate whether there is pollution in this Water Supply area that can cause the public health risk.

2-The coordinate of 2 sides of the location:

- SG2a: N 10^0 59’ 8,16” E 106^0 37’ 20,47”
- SG2b: N 10^0 59’ 5,36” E 106^0 37’14,02”

3-Biological replications:

- The distance between these sites is about 50 – 100 meters, so that we extract a total of 6 samples in this location.

A.3. Location SG3:

1-Description:

Samples were collected at the shore of the 2 sides of the river, under the bridge, near and downstream of the Binh Phuoc bridges poles (**Fig. A.4**). There are several wastewater discharge drains of the surrounding urban area. We chose this site to take the sample to investigate the pollution affected by the industrial parks and densely populated area from Binh Duong Province.

2-The coordinate of 2 sides of the location:

SG3a: N 10^0 51' 42,63" E 106^0 43' 4,74"

SG3b: N 10^0 51' 39,37" E 106^0 42' 58,71"

3-Biological replications:

The distance between the sites b1, b2 is about 50 meters so that we extracted the total DNA from all sites.

A.4. Tan Thuan Bridge (SG6):

1-Description:

We took the samples in this location because it lies near the Tan Thuan Industrial Park, river ports and densely populated of district 7, HCMC. The samples were collected at the intersection of Kenh Te canal and the SaiGon river.

2-The coordinate of 2 sides of the location:

- SG6a: N 10^0 45' 38,26" E 106^0 43' 21,89"
- SG6b: N 10^0 45' 25,40" E 106^0 43' 15,05"

3-Biological replications:

The distance between the sites a1, a2, a3 is not very far (about 1 meter), so that we extracted total DNA from the a1 and a2 samples.

A.5. Location SG5:

1-Description:

Samples were taken at Mui Den Do T-Junction. This is the intersection of the SaiGon river and the DongNai river so that the water flow is intense and there is less sediment in this location compared to the other locations (specially the location near Phu My Bridge).

2-The coordinate of 2 sides of the location:

- SG5a: N 10^0 44' 15,38" E 106^0 46' 18,93"
- SG5b: N 10^0 44' 52,18" E 106^0 45' 47,57"

3-Biological replications:

The distance between these sites is about 50 – 100 meters, so that we extracted DNA of total 4 sediment samples.

Figure A.1. Photos of the surrounding area of SG5 location (one side of the river).

A.6. Locaton SG4:

1-Description:

Samples were collected far from Cat Lai ferry station about 50 m, toward downstream direction of DongNai river.

We took sediment samples at the shores belong 2 sides of the river under the bridge. This is the main discharge water of the surrounding densely populated area.

2-The coordinate of 2 sides of the location:

- SG4a: N 10^0 45’ 6,59” E 106^0 47’ 24,46”
- SG4b: N 10^0 45’ 25,49” E 106^0 47’ 15,44”

3-Biological replications:

The distance between the sites b1, b2, b3 is about 50 meters, so that we extracted total DNA from all the samples.


Figure A.2. Photos of the surrounding areas of SG4 location (one side of the river).

A.7. Location DN1: `

1-Description:

The samples were taken at the shore of the 2 sides of the river, near the Hoa An Water Pumping Station. Samples were collected at 2 sides of the river and located 50m upstream the Hoa An bridge. The sample at the location DN1b (belonging to Binh Duong province) is difficult to collect because there was construction activity near Hoa An pumping station.

2-The coordinate of 2 sides of the location:

- DN1a: N 10^0 57’ 0,67” E 106^0 48’ 21,95”
- DN1b: N 10^0 56’ 43,09” E 106^0 48’ 16,14”

3-Biological replications:

The distance between the sites a1, a2, a3 is not very far (about 1 meter), so that we extract total DNA from a1 and a2 samples.


Figure A.3. Photos of the surrounding areas of DN1 location.

A.8. Location DN2:

1-Description:

Samples were collected at the 2 sides of the river upstream the Dong Nai bridge. The sediment samples DN2a were collected near the sewage system of the Bien Hoa I industrial park. This location is affected by the industrial activities of Bien Hoa I and II industrial parks. However, all the wastewater from these two industrial parks has been treated by the waste water treatment plant. The sediment samples DN2b located near the Tan Van canal flow toward to the Dong Nai river.

From Hoa An Bridge, where we took sample DN1, to Mui Den Do T-Join, where we took sample SG5, there are many Industral Parks of Dong Nai Province and Bien Hoa City.

2-The coordinate of 2 sides of the location:

- DN2a: N 10^0 54’ 28,39” E 106^0 50’ 29,20”
- DN2b: N 10^0 54’ 15,11” E 106^0 50’ 12,21”

3-Biological replications:

The distance between the sites a1, a2, a3 and b1, b2, b3 are about 50 meters, so that we extract total DNA from all the samples.

A.9. Location RF1:

1-Description:

Samples were collected at Ben Suc bridge, belonging to the border area between HCMC and Binh Duong province. The samples were collected at two sides of the river and far from Ben Suc bridge about 100 meter.

2-The coordinate of 2 sides of the location:

- Ref1a: N 11^0 9’ 19,44” E 106^0 27’ 4,48”
- Ref1b: N 11^0 9’ 18,99” E 106^0 27’ 6,92”

3-Biological replications:

The distance between the sites a1, a2, a3 and b1, b2, b3 is not very far (about 1 meter), so that we extracted total DNA from the a1, a2 and b1, b2 sites.

A.10. Location RF2 (Thien Tan Water Supply Factory):

1-Description:

Samples were taken 4 km downstream of the Song Be and Dong Nai river T-junction and nearby the sand transporting station. The sediment has brick-red color, with a lot of sand particles due to the natural characteristics of the river and because it is near the sand transportating station.

2-The coordinate of 2 sides of the location:

- Ref2a: N 11^0 4’ 7,04” E 106^0 57’ 2,93”
- Ref2b: N 11^0 4’ 10,50” E 106^0 56’ 59,90”

3-Biological replications:

The distance between the sites a1, a2, a3 and b1, b2, b3 are about 50 meters, so that we extracted total DNA from all the samples.

Figure A.4. Photos of the surrounding areas of RF2 location.

A.11. Location SG8:

1-Description:

Samples supposed to be collected at two sides of the canals, one is upstream and one is downstream of Cau Nho Bridge on To Ngoc Van Street (also called Highway 43) near the Binh Chieu Industrial Park. This is the location which receives sewage from Binh Chieu, Linh Trung 2 industrial parks, and surrounding urban area surrounding. However, the width of the canal is very small so that the team decided to take just one sample for this location. The canal has been heavily contaiminated. The sediment has a black color and strong odor (**Fig. A.20**).

2-The coordinate of 2 sides of the location:

- SG8a: N 10^0 53’ 15,12” E 106^0 43’ 44,03”
- SG8b: N 10^0 53’ 14,95” E 106^0 43’ 442,99

3-Biological replications: Just one sample was taken which are called SG8a1.

Figure A.5. Photos of the surrounding area of SG8 location. Right photo is the sediment sample.

A.12. Location SG9:

1-Description:

Samples were collected 200m from Kenh Te bridge and located at the Police Station of Waterway Traffic, District 4, HCMC. The location receives sewage from the surrounding urban area and has been heavily polluted. The sediment has the dark black color and strong odor.

Due to the transportation difficulties, we collected one sample from the left side of this location (SG9a).

2-The coordinate of 2 sides of the location:

- SG9a: N 10^0 45’ 9,76” E 106^0 42’ 0,45”

3-Biological replications: Only one sample was taken which is called SG9a1.

Annex 2: Distribution of industrial parks (IPs) around sampling locations.

A.2.1. Method of mapping the IPs located in the studied area of SG-DN river system:

The method of mapping the IPs located in the studied area used 3 steps:

1- From the list of IPs of each province and HCMC in the year 2012, each address was collected.

2-Posted on Google Map to find the location of each IP.

3-Map the point of the location on the Map of the SG-DN river (**Fig. A.22**).

A.2.2. The list of Industrial Parks (IPs):

A.2.2.1. In Dong Nai province:

The list of Industrial Parks (IPs) located in the Dong Nai province was downloaded from the website of Dong Nai Industrial Zones Authority on 12th March 2012.

- Vietnamese: BAN QUAN LY CAC KHU CONG NGHIEP DONG NAI
- Link: http://diza.dongnai.gov.vn/Pages/kcn.aspx

There were a total of 30 IPs present in Dong Nai province with an area ranging from 43 ha to 529 ha and were established from 1994 to 2010. The total area was 9572 ha.

A.2.2.2. In Binh Duong province:

The list of Industrial Parks (IPs) located in Binh Duong province was downloaded from the website of Binh Duong Industrial Zones Authority on 04th March 2012.

-Vietnamese: BAN QUAN LY CAC KHU CONG NGHIEP BINH DUONG.

-Link: http://kcn.binhduong.gov.vn/Lists/ThongTinCacKCN/TongQuat.aspx?PageIndex=0

There were a total of 24 IPs present in the Binh Duong province with an area ranging from 16.5 ha to 997.7 ha and were established until 2012. The total area was 6198.9 ha.

A.2.2.3. In HCMC:

The list of Industrial Parks (IPs) located in the HCMC was downloaded from the website of HCMC Industrial Zones Authority on 12th March 2012.

-Vietnamese: BAN QUAN LY CAC KHU CHE XUAT & KHU CONG NGHIEP TP.HCM

-Link: http://www.hepza.hochiminhcity.gov.vn/web/guest/kcn_kcx-tphcm/bang-gia-dat

There were a total of 17 IPs present in the HCMC with an area ranging from 62 ha to 597 ha and were established until 2012. The total area was 3583,2 ha.

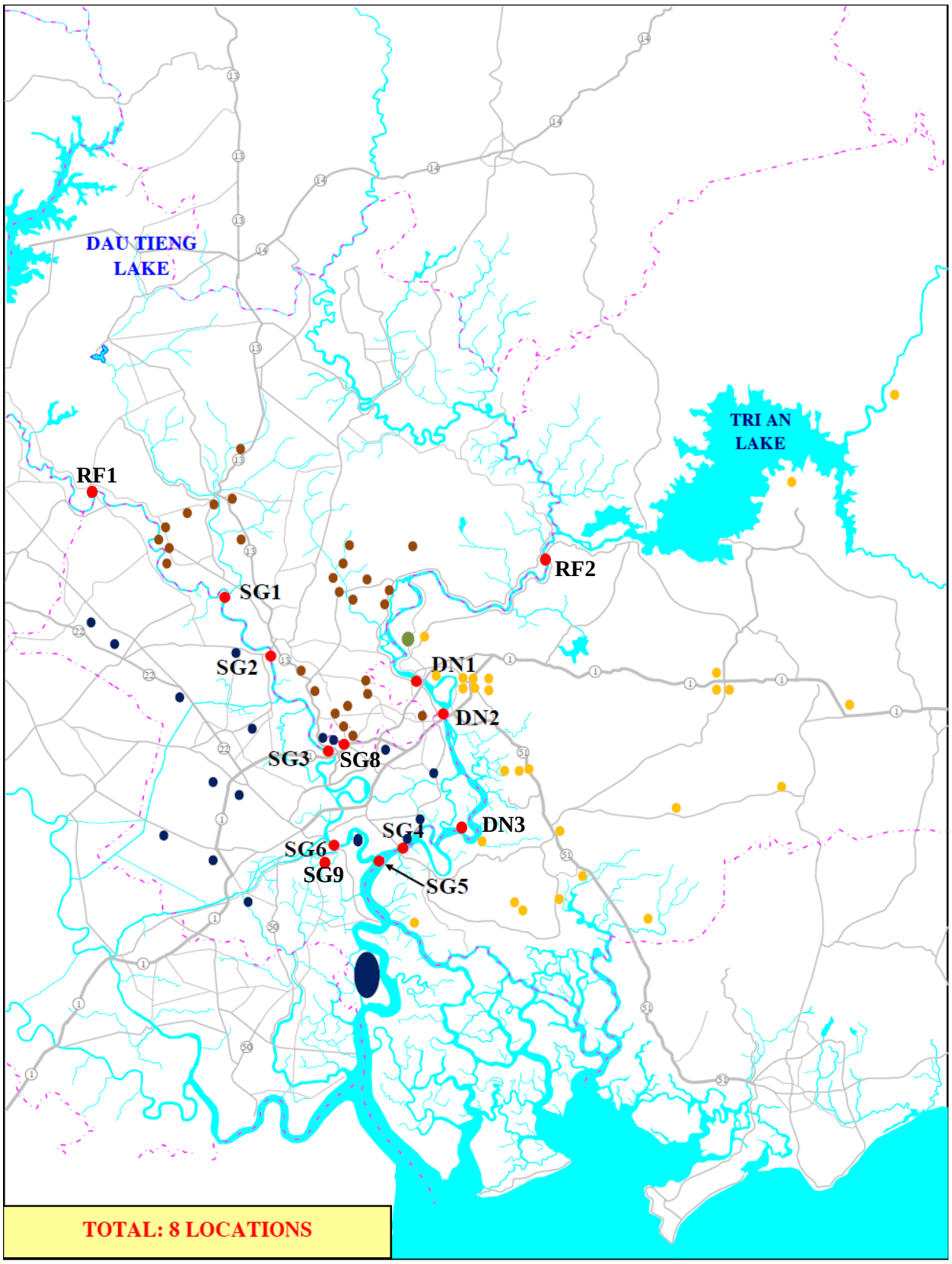


Figure A.6. Map of 13 locations and the distribution of Industrial Parks on the SaiGon-DongNai rivers.

Annex 3: Three runs of 454 pyrosequencing with Mids and Sample ID

Table A.1: Three runs of 454 pyrosequencing GS Junior System, each run has 15 samples (14 sediment samples and 1 control sample)

Number	Mid sequences	Sample ID					
		First run		Second run		Third run	
1	ACGAGTGCGT	I-1	RF2_a1	II-1	RF1_a1	III-1	DN3_a1
2	ACGCTCGACA	I-2	RF2_a2	II-2	RF1_a2	III-2	DN3_a2
3	AGACGCACTC	I-3	RF2_b1	II-3	RF1_b1	III-3	DN3_b1
4	AGCACTGTAG	I-4	RF2_b2	II-4	RF1_b2	III-4	SG4_a1
5	ATCAGACACG	I-5	DN1_a1	II-5	SG1_a1	III-5	SG4_b1
6	ATATCGCGAG	I-6	DN1_a2	II-6	SG1_a2	III-6	SG4_b2
7	CTCGCGTGTC	I-7	DN1_b1	II-7	SG2_a1	III-7	SG4_b3
8	TAGCGCATAC	I-8	DN2_a1	II-8	SG2_a2	III-8	SG5_a1
9	TCTCTATGCG	I-9	DN2_a2	II-9	SG2_a3	III-9	SG5_a2
10	TATAGCGCAC	I-10	DN2_a3	II-10	SG2_b1	III-10	SG5_b1
11	CATAGTAGTG	I-11	DN2_b1	II-11	SG2_b2	III-11	SG5_b2
12	CGAGAGATAC	I-12	DN2_b2	II-12	SG3_a1	III-12	SG6_a1
13	TCACGTACTA	I-13	DN2_b3	II-13	SG3_b1	III-13	SG6_a2
14	CGTCTAGTAC	I-14	SG2_b3	II-14	SG8_a1	III-14	SG9_a1
15	ACGACTACAG	I-15	Control 1	II-15	Control 2	III_15	Control 3

Note:

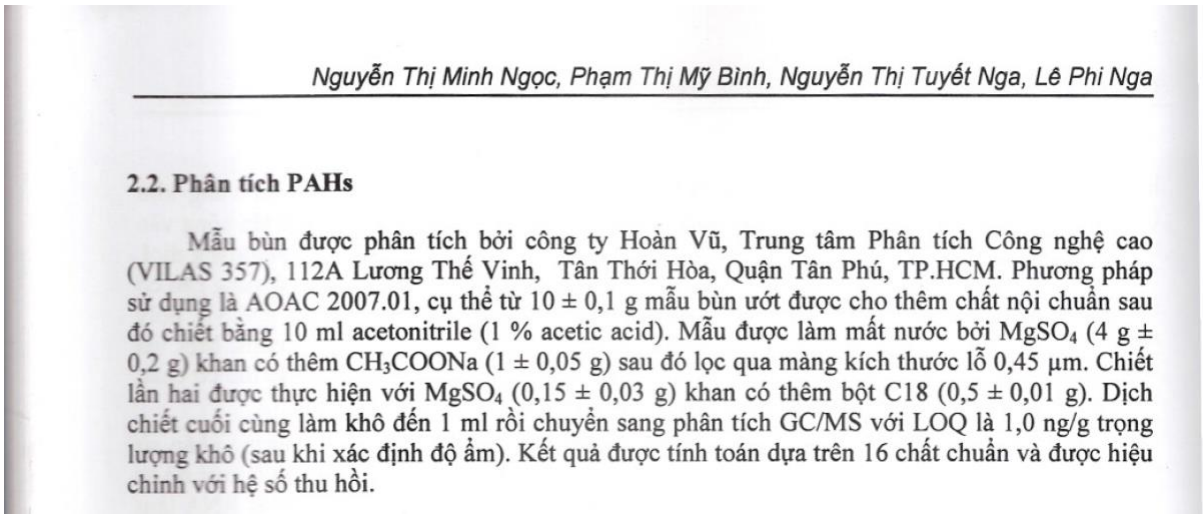
Control 1: *E. coli* K-12 MG 1655 DNA extraction.

Control 2: *E. coli* K-12 MG 1655 & Deinococcus radiodurans DNA extractions that were PCR amplified separately, and then mixed to send for pyrosequencing with other 14 samples

Control 3: *E. coli* K12 MG 1655 & Deinococcus radiodurans DNA extraction that were PCR amplified together, and then send for pyrosequencing with other 14 samples.

Annex 4: Translation of PAHs method from National Article (published by colabroration Prof. Le Phi Nga, University of Natural Science, Ho Chi Minh City, Vietnam)

The method for analysing PAHs is translated from the Vietnamese Article. The PAHs method is in page 264, 1st paragraph of the article (Figure A.15)



Nguyễn Thị Minh Ngọc, Phạm Thị Mỹ Bình, Nguyễn Thị Tuyết Nga, Lê Phi Nga

2.2. Phân tích PAHs

Mẫu bùn được phân tích bởi công ty Hoàn Vũ, Trung tâm Phân tích Công nghệ cao (VILAS 357), 112A Lương Thế Vinh, Tân Thới Hòa, Quận Tân Phú, TP.HCM. Phương pháp sử dụng là AOAC 2007.01, cụ thể từ 10 ± 0,1 g mẫu bùn ướt được cho thêm chất nội chuẩn sau đó chiết bằng 10 ml acetonitrile (1 % acetic acid). Mẫu được làm mất nước bởi $MgSO_4$ (4 g ± 0,2 g) khan có thêm CH_3COONa (1 ± 0,05 g) sau đó lọc qua màng kích thước lỗ 0,45 µm. Chiết lần hai được thực hiện với $MgSO_4$ (0,15 ± 0,03 g) khan có thêm bột C18 (0,5 ± 0,01 g). Dịch chiết cuối cùng làm khô đến 1 ml rồi chuyển sang phân tích GC/MS với LOQ là 1,0 ng/g trọng lượng khô (sau khi xác định độ ẩm). Kết quả được tính toán dựa trên 16 chất chuẩn và được hiệu chỉnh với hệ số thu hồi.

Figure A.7. Image of PAHs method from National Vietnamese Article (**236**).

Translation:

Sediment samples were analyzed by HoanVu Laboratory, Scientist Technologies Company Limited (VILAS 357) (1,2), address 112 Luong The Vinh, Tan Thoi Hoa, Tan Phu District, HCM city, Vietnam. Method used is AOAC 2007.01 (3) with following steps (4):

(a) starting with 10 ± 0,1 g wet sediment

(b) adding with spiking solution

(c) adding 10 mL of acetonitril (1% acetic acid) , $MgSO_4$ (4 ± 0,2g), CH_3COONa (1 ± 0.05 g) then filtering through 0.45 µm filter.

(d) The second extraction was performed with $MgSO_4$ (0,15 ± 0,03 g) and C18 (0,5 ± 0.01 g).

(e) the extract were dried to 1.0 mL

(f) 1.0 mL extract were sent for analysis by GC/MS (gas chromatography–mass spectrometry) with LOQ (limit of quantification) is 1,0 ng/g dry weight (after identify the humidity index).

(g) The results were calculated based on 16 standard compounds.

Note:

(1) About Hoan Vu company. Website: http://www.hoanvulab.com/

HoanVu Scientific Ltd. was established in April 2007 under permit number: 4104000709 of the Ministry of Planning and Investment. Hoan Vu Scientific is an independent private laboratory. It was established with 100% foreign investment by Mr. Henry Bui, a Vietnamese-American, who has been working in the mass spectrometer field in the United States for the past 25 years. He specializes in the service and support of existing mass spectrometers. Mr. Bui is the founder of Hoan Vu Scientific and Cal-Tech Scientific, Inc. in California, USA. Hoan Vu is holding a valid Certificate ISO/IEC 17025:2005 with VILAS 357 (2).

Website:http://www.hoanvulab.com/gi7899ithi7879uho224nv361.html

(2) VILAS 357

VILAS is the certificated laboratory system in Vietnam originated from Bureau of Accreditation Vietnam (BoA). Website: http://www.boa.gov.vn/

(3) AOAC 2007.01 method:

AOAC 2007.01 method for PAHs for soil samples is called: "Analysis of Polycyclic Aromatic Hydrocarbons in Soil with Agilent SampliQ QuEChERS AOAC Kit and HPLC-FLD from Agilent Technologies (Interchim)." However, the AOAC 2007.01 method in the article is quite different from the standard AOAC 2007.01.

(4) About AOAC:

Association of Official Analytical Chemists (AOAC) is a non-profit scientific association with headquarters in Rockville, Maryland, USA. It publishes standardized, chemical analysis methods designed to increase confidence in results of chemical and microbiologic analyses. Since 1884, AOAC INTERNATIONAL has ensured the ability of analytical scientists to have confidence in their results through the adoption of methods as AOAC®Official Methods℠.Website: http://www.aoac.org/

(5) The method in the article is more similar to the study below with a few differences (underlined):

PAHs method in the study: Brondi SHG, De MacEdo AN, Vicente GHL, Nogueira ARA. Evaluation of the QuEChERS method and gas chromatography-mass spectrometry for the analysis pesticide residues in water and sediment. Bull Environ Contam Toxicol. 2011; 86(1):18–22.

Our method	The study
(a) starting with 10 ± 0,1 g wet sediment	(a) placing a sample of 10 g of water or dry sediment into a centrifuge tube;
(b) adding with spiking solution	(b) adding atrazine, fipronil and endosulfan in the required concentrations;
(c) adding 10 mL of acetonitril (MeCN) (1% acetic acid), $MgSO_4$ (4 ± 0,2g), CH_3 COONa (5) (1 ± 0.05 g) then filtering through 0.45 μm filter.	(c) adding 10 mL of MeCN, 4 g of $MgSO_4$ and 1 g of NaCl (5) in each tube, and centrifuging it at 3,000 rpm for 1 min;
(d) The second extraction were done with $MgSO_4$ (0,15 ± 0,03 g) and C18 (0,5 ± 0.01 g).	(d) transferring 5 mL of MeCN extract to a commercial SPE cartridge containing 330 mg PSA, 330 mg C18 and a 1 cm layer of $MgSO_4$ activated with 3 mL of MeCN.
(e) the extract were dried to 1.0 mL	(e) the extract was passed and collected
(f) 1.0 mL extract were sent for analysis by GC/MS (gas chromatography–mass spectrometry) with LOQ (limit of quantification) is 1,0 ng/g dry weight (after identify the humidity index).	(f) 1.0 mL of the extract was transferred to an autosampler vial (Shimadzu AOC-20i autoinjector – Kyoto, Japan) for analysis by GC–MS. The volume analyzed was 1 μL
(g) The results were calculated base on 16 standard compounds	

Note:

(i) CH_3 COONa and NaCl has the same function since they donor the ion Na^+

(6) Detection limit for PAHs using GC-MS

Detection thresold 17 PAH standard compounds is presented below **(358):**

Table 2. Response factor, detection limits, recoveries of check standards, and relative percent differences of sample duplicates for individual PAHs in this study.

Compound	Response factor (RF) (n = 5)		Detection limits	Check analysis (n = 7)	Duplication analysis (n = 7)
	Average ± SD [a]	RSD [a] (%)	DL (ng/g)	R [a] (%)	RPD [a] (%)
Naphthalene	2.08 ± 0.17	8.4	2.9	122 ± 12	8.7 ± 4.6
Acenaphthylene	1.85 ± 0.11	5.9	1.4	87 ± 6	11.8 ± 3.8
Acenaphthene	1.14 ± 0.05	4.4	1.9	107 ± 9	9.2 ± 5.0
Fluorene	0.86 ± 0.01	1.1	0.6	98 ± 3	11.2 ± 5.4
Phenanthrene	1.09 ± 0.14	12.9	2.3	105 ± 9	10.7 ± 3.3
Anthracene	1.28 ± 0.10	7.6	2.0	89 ± 8	7.0 ± 6.0
Fluoranthene	1.17 ± 0.06	4.8	1.8	90 ± 9	8.9 ± 2.4
Pyrene	1.22 ± 0.07	5.9	1.7	90 ± 8	12.3 ± 2.3
Benzo[a]anthracene	0.97 ± 0.11	11.6	2.2	105 ± 9	13.3 ± 3.6
Chrysene	1.47 ± 0.20	13.7	2.2	105 ± 8	11.9 ± 7.2
Benzo[b]fluoranthene	0.79 ± 0.09	11.9	3.5	120 ± 16	12.3 ± 4.3
Benzo[k]fluoranthene	1.39 ± 0.14	10.2	3.0	107 ± 14	10.8 ± 2.1
Benzo[a]pyrene	0.78 ± 0.03	4.2	3.5	91 ± 16	12.3 ± 4.6
Indeno[1,2,3-cd]pyrene	0.64 ± 0.06	9.3	4.4	103 ± 18	12.4 ± 6.4
Dibenz[a,h]anthracene	0.41 ± 0.06	14.1	5.4	112 ± 14	11.2 ± 4.9
Benzo[g,h,i]perylene	0.72 ± 0.07	9.9	5.3	128 ± 4	10.1 ± 7.4
2-Fluorobiphenyl (SS1)	1.52 ± 0.09	5.86	-	102 ± 7	7.5 ± 2.5
4-Terphenyl-d_{14} (SS2)	1.41 ± 0.10	7.09	-	107 ± 18	9.2 ± 4.0

[a] SD: standard deviation; RSD: Relative standard deviation; R: Recoveries; RPD: Relative percent differences.

The Detection Limit of 17 PAH standard compounds is approximately 1.0 $ng.g^{-1}$.

REFERENCES:

1. The Water Cycle. The USGS (United States Geological Survey) Water Science School. Retrieved from URL: http://water.usgs.gov/edu/watercycle.html

2. Saline water in the United States. The USGS Water Science School. Retrieved from URL: http://water.usgs.gov/edu/salineuses.html.

3. River. The USGS Water Science School. Retrieved from URL: http://water.usgs.gov/edu/earthrivers.html

4. Corfield J. (2013). Historical Dictionary of Ho Chi Minh City. London, Anthem Press. Page 272.

5. Sustainable Groundwater Management in Asian Cities. Fresh Water Resources Management Project. Institute for Global Environmental Strategies (IGES), 2007. Chapter 3-3, page: 69-71. Retrieved from URL: http://enviroscope.iges.or.jp/modules/envirolib/upload/981/attach/00_complete_report.pdf

6. World Regional Geography: People, Places and Globalization.Libraries. University of Minnesota. Retrieved from URL: http://open.lib.umn.edu/worldgeography/chapter/11-2-the-mainland-countries/

7. Campodia Physical Map. Free World Map. Retrieved from URL: http://www.freeworldmaps.net/asia/cambodia/map.html

8. Nguyễn Ty Niên. Quản lý tổng hợp tài nguyên nước lưu vực sông Đồng Nai- Một yêu cầu cấp bách. VNCOLD. Retrieved from URL: http://www.vncold.vn/Web/Content.aspx?distid=1051

9. Population Total. The World Bank. Retrieved from URL: http://data.worldbank.org/indicator/SP.POP.TOTL

10. Vietnam Population. Worldometer. Retrieved from URL: http://www.worldometers.info/world-population/vietnam-population/

11. Population and population density in 2011 by district. Statistical office in Ho Chi Minh City. P. 23. Retrieved from URL: http://www.pso.hochiminhcity.gov.vn/c/document_library/get_file?uuid=bb171c42-6326-4523-9336-01677b457b13&groupId=18

12. Demographia World Urban Areas (World Agglomerations), 7th Annual Edition, April 2011. Page 13. Retrieved from URL: http://www.laschools.net/cms/lib07/NM01000458/Centricity/Domain/575/db-worldua.pdf

13. Statistical Yearbook of Vietnam 2011. General Statistics Office of Vietnam. Part 16: Area, population and population density in 2011 by province ; page 60. Retrieved from URL: http://www.gso.gov.vn/default.aspx?tabid=512&idmid=5&ItemID=12574

14. Half of Vietnam's industrial sewage dumped into rivers untreated. Thanhnien News. October, 2013. Retrieved from URL: http://www.thanhniennews.com/society/half-of-vietnams-industrial-sewage-dumped-into-rivers-untreated-956.html

15. Pollution worsens in Dong Nai River. Viet Nam News. July, 2013. Retrieved from URL: http://vietnamnews.vn/environment/241593/pollution-worsens-in-dong-nai-river.html

16. Hays J., Water Pollution in China. Facts and Details. 2008, last updated April 2014. Retrieved from URL: http://factsanddetails.com/china/cat10/sub66/item391.html

17. Li X, Liu L, Wang Y, Luo G, Chen X, Yang X. Integrated assessment of heavy metal contamination in sediments from a coastal industrial basin, NE China. PLoS One. 2012; 7(6):e39690.

18. Central Pollution Control Board, India, Annual Report 2008-2009. Central Pollution Control Board, Ministry of Environment & Forests, Government of India. 2009. Retrieved from URL: http://cpcb.nic.in/upload/AnnualReports/AnnualReport_37_ANNUAL_REPORT-08-09.pdf

19. Kansas State University, Freshwater Pollution Costs US At Least $4.3 Billion A Year. ScienceDaily. November, 2008. Retrieved from URL: http://www.sciencedaily.com/releases/2008/11/081112124418.htm

20. Lao W, Tsukada D, Greenstein DJ, Bay SM, Maruya K a. Analysis, occurrence, and toxic potential of pyrethroids, and fipronil in sediments from an urban estuary. Environ Toxicol Chem. 2010; 29(4):843–51.

21. Guevara S. Toxic Spill in Sonora The tip of the iceberg. Greenpeace. September, 2014. Retrieved from URL: http://www.greenpeace.org/usa/toxic-spill-sonora-tip-iceberg/

22. Junior JC, Neves V, Amendola P, Rocha RM, Costa JJG Da. Water Statistics in Brazil: an Overview. Instituto Brasileiro de Geografia e Estatística. 2005.

23. Homad-Hamam G. Chile Deals With Increased Region V Forest Fires. Santiago Times. December 2009. Retrieved from URL: http://engineering.columbia.edu/files/engineering/design-water-resource04.pdf

24. Abouzaid H, Echihabi L. Drinking water quality and monitoring in north Africa: the Moroccan experience. Science Total Environ. 1995; 171(1-3):29–34.

25. Barakat A, Baghdadi M. El. Assessment of Heavy Metal in Surface Sediments of Day River at Beni-Mellal Region, Morocco. Research Journal of Environmental and Earth Sciences. 2012; 4(8):797–806.

26. Tap water 'polluted' for 1.5 million in France. The Local Fr. February 2014. Retrieved from URL: http://www.thelocal.fr/20140226/15-million-french

27. Sánchez-Chóliz J, Duarte R. Water pollution in the Spanish economy: Analysis of sensitivity to production and environmental constraints. Ecological Economics. 2005; 53:325–38.

28. Gilbert N. Europe sounds alarm over freshwater pollution. Nature News. March 2015. Retrieved from URL: http://www.nature.com/news/europe-sounds-alarm-over-freshwater-pollution-1.17021

29. Rungsuk P. Thailand's Rivers Polluted by Factory and Residential Waste. Environment News Service. September, 2011. Retrieved from URL: http://ens-newswire.com/2011/09/27/thailands-rivers-polluted-by-factory-and-residential-waste/

30. Thongraar W., Chaluay Musika W.W. and AM. Heavy metals contamination in sediments along the Eastern Coast of the Gulf of Thailand. Environment Asia. 2008; 1:37–45.

31. Sudaryanto A., Isobe T., Takahashi S., Tanabe S. Assessment of persistent organic pollutants in sediments from Lower Mekong River Basin. Chemosphere. 2011; 82(5):679–86.

32. China says water pollution so severe that cities could lack safe supplies. Chinadaily. June, 2005. Retrieved from URL: http://www.chinadaily.com.cn/english/doc/2005-06/07/content_449451.htm

33. UNESCO (2003). Water for People Water for Life. WWAP (World Water Assessment Programme). The United Nations World Water Development Report. 2003. Retrieved from URL: http://www.un.org/esa/sustdev/publications/WWDR_english_129556e.pdf

34. Singh Chabba AP. Water-Borne Diseases in India. Reset. May 2013. Retrieved from URL: https://en.reset.org/blog/water-borne-diseases-india

35. Kahn J, Yardley J. As China Roars, Pollution Reaches Deadly Extremes. New York Times. August 2007. Retrieved from URL: http://www.nytimes.com/2007/08/26/world/asia/26china.html?_r=0

36. Hense BA, Jaser W, Welzl G, Pfister G, Wöhler-Moorhoff GF, Schramm K-W. Impact of 17alpha-ethinylestradiol on the plankton in freshwater microcosms--II:

responses of phytoplankton and the interrelation within the ecosystem. Ecotoxicol Environ Saf. 2008; 69(3): 453–65.

37. Schramm KW, Jaser W, Welzl G, Pfister G, Wöhler-Moorhoff GF, Hense B a. Impact of 17α-ethinylestradiol on the plankton in freshwater microcosms-I: Response of zooplankton and abiotic variables. Ecotoxicol Environ Saf. 2008; 69:437–52.

38. Dutta S.. India is already facing a water crisis—and it is only going to get worse. India Quartz. March, 2015.
Retrieved from URL: http://qz.com/353707/india-is-already-facing-a-water-crisis-and-it-is-only-going-to-get-worse/

39. Hogan C. M. Water pollution. The encyclopedia of Earth. November, 2014.
Retrieved from URL: http://www.eoearth.org/view/article/156920/

40. Jana, B. Kr., Majumder M. Impact of climate change on natural resource management. Environ Science. 2010; 207.

41. Mohammed AS, Kapri A, Goel R. Biomanagement of metal-contaminated Soils. Environ Science. 2011; (1): 1-28.

42. Duruibe JO, Ogwuegbu MOC, Egwurugwu JN. Heavy metal pollution and human biotoxic effects. Int J Phys Sci. 2007; 2(5):112–8.

43. Dikshith TSS. Hazardous Chemicals: Safety Management and Global Regulations. Florida. CRC Press. 2013; p 494.

44. Chamy R, Rosenkranz F. Biodegradation - life of science. Environ Technol. 2013; (11): p1.

45. Hobson PN. Bioconversion of waste materials to industrial products. Elsevier Applied Biotechnology Series. 1998; page 317.

46. Bhateria R, Jain D. Water quality assessment of lake water: a review. Sustain Water Resour Manag. 2016; 2(2):161–73.

47. Emanuel E., Arnold J. Bloom. Mineral nutrition of plants mineral nutrition of plants: principles and perspectives. BioScience. 1972.

48. Heldt H-W, Piechulla B. Plant Biochemistry. London. Academic Press. 2011 ; page 438-466.

49. Sumpter JP. Protecting aquatic organisms from chemicals: the harsh realities. Philos Trans A Math Phys Eng Sci. 2009; 367(1904):3877–94.

50. Stanton AT, Fletcher W. Melioidosis, a new disease of the tropics. Far Eastern Association of Tropical Medicine. Transactions of the Fourth Congress. 1921; (2):196–8.

51. Saravua K, Vishwanatha S, Kumar RS, Barkur AS, Varghese GK, Mukhyopadhyay C, et al. Melioidosis: A case series from south India. Trans R Soc Trop Med Hyg. 2008; 102:S18–20.

52. Suputtamongkol Y, Chaowagal W, Chetchotisakd P, Lertpatanasuwun N, Intaranongpai S, Ruchutrakool T, et al. Risk factors for melioidosis and bacteremic melioidosis. Clin Infect Dis.1999; 29:408–13.

53. Cheng AC, Jacups SP, Ward L, Currie BJ. Melioidosis and Aboriginal seasons in northern Australia. Trans R Soc Trop Med Hyg. 2008; 102 Suppl :S26–9.

54. Suputtamongkol Y, Chaowagul W, Chetchotisakd P, Lertpatanasuwun N, Intaranongpai S, Ruchutrakool T, et al. Risk factors for melioidosis and bacteremic melioidosis. Clin Infect Dis. 1999; 29(2):408–13.

55. Neliyathodi S, Thazhathethil AN, Pallivalappil L, Balakrishnan D. Pleuropulmonary melioidosis with osteomyelitis rib. Lung India. 2015; 32(1):67–9.

56. Giannella RA. Salmonella. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 21.

57. Evans DJ Jr., Evans DG. Escherichia coli in diarrheal disease. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 25.

58. Finkelstein RA. Cholera, Vibrio cholerae O1 and O139, and Other Pathogenic Vibrios. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 24.

59. Iglewski BH. Pseudomonas. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 27.

60. Peterson JW. Bacterial Pathogenesis. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 7.

61. Meyer EA. Other Intestinal Protozoa and Trichomonas Vaginalis. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 80.

62. Blacklow NR. Norwalk Virus and Other Caliciviruses. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 65.

63. Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ. Heavy metal toxicity and the environment. EXS. 2012; 101:133–64.

64. Rathoure, Ashok K. Toxicity and Waste Management Using Bioremediation. India.

IGI Global Publisher. 2015; Chapter 2, page 35.

65. United Nations Environmental Protection/Global Program of Action (2004). Why The Marine Environment Needs Protection From Heavy Metals, Heavy Metals 2004, UNEP/GPA Coordination Office.

66. Jarup L. Hazards of heavy metal contamination. Br Med Bull. 2003; 68:167–8.

67. Jaishankar M, Tseten T, Anbalagan N, Mathew BB, Beeregowda KN. Toxicity, mechanism and health effects of some heavy metals. Interdiscip Toxicol. 2014; 7(2):60–72.

68. Vo LP. Water resource management in Ho Chi Minh City, Vietnam - An overview. Science and Technology development. 2009; p. 51–63.
Retrieved from
URL:http://www.vjol.info/index.php/JSTD/article/viewFile/2123/2649

69. Pollution soon to render Dong Nai river unsuable. Thanhnien News. December 2009. Retrieved from URL: http://www.thanhniennews.com/society/pollution-soon-to-render-dong-nai-river-unusable-17836.html

70. Duc Hiep Nguyen TTP. Water resources and environment in an around Ho Chi Minh City, Vietnam. Electron Green J. 2003; 1(19).
Retrieved from URL: https://escholarship.org/uc/item/2tk6z0xg

71. Pollution levels in Dong Nai river on red alert. Vietnam Plus. April 2010.
Retrieved from URL: http://en.vietnamplus.vn/Home/Pollution-levels-in-Dong-Nai-River-on-red-alert/20104/7908.vnplus

72. Vietnam: Even bottled water unsafe. IRIN. April 2009.
Retrieved from URL: http://www.irinnews.org/report/83965/vietnam-even-bottled-water-unsafe

73. Binh An Water Corporation Limited: Always Supplying High-Quality Potable Water. Vietnam Chamber of Commerce and Industry. February, 2016 (last updated).
Retrieved from URL: http://vccinews.com/news_detail.asp?news_id=4226

74. ENVIRONMENT-VIETNAM: River Pollution Scandal a Wake-up Call. Inter Press Service. December 2008.
Retrieved from URL: http://www.ipsnews.net/2008/12/environment-vietnam-river-pollution-scandal-a-wake-up-call/

75. Boycott fear forces river polluter payout. ThanhNienNews. August 2010.
Retrieved from URL: http://www.thanhniennews.com/society/boycott-fear-forces-river-polluter-payout-15358.html

76. Search on Google Map with the address of Vedan Company on Thi Vai river. Address of Vedan company in Vietnamese: "Quốc Lộ 51, Ấp 1A, Xã Phước Thái, Huyện Long Thành, Phước Thái, Long Thành, Đồng Nai, Vietnam." **Note:** Đồng

Nai means Dong Nai province (one of 10 provinces located on the Saigon-Dongnai river system).

77. Vietnam to clean river with $4 mil mangroves project. ThanhNienNews. March 2014. Retrieved from URL: http://www.thanhniennews.com/education-youth/vietnam-to-clean-river-with-4-mil-mangroves-project-24608.html

78. LE VIETNAM : BONJOUR LA POLLUTION. Blouge sur l'Asie du Sud-Est. November 2014. Retrieved from URL: http://redtac.org/asiedusudest/2014/11/16/le-vietnam-bonjour-la-pollution/

79. Prilop K, Quan NH, Lorenz M, Huyen Le, Hien LT, Meon G. Intergrated water quality monitoring of the Thi Vai river : an assessment of historical and current situation. "Green growth, climate change and protection of the coastal environment" 17th + 18th of June, Ho Chi Minh City, Vietnam . 4th VNU – HCM International Conference for Environment and Natural Resources ICENR 2014.

80. Pink M R. Water Right in Southeast Asia and India Book. Palgrave Macmillan US. November 2015. p. 211.

81. Vedan admits to polluting parts of Thi Vai River. Viet Nam News. December 2009. Retrieved from URL: http://vietnamnews.vn/print/194958/vedan-admits-to-polluting-parts-of-thi-vai-river.html

82. Viet Nam Water and Sanitation Sector Assessment Strategy and Roadmap. Asian Development Bank. June 2010, page: 2-14.
Retrieved from URL:: http://www.wastewater-vietnam.org/images/201006.ADB.VN%20WaterSanitation%20AssessmentStrategyRoadmap.pdf

83. Tamaki H, Sekiguchi Y, Hanada S, Nakamura K, Nomura N, Matsumura M, et al. Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. Appl Environ Microbiol. 2005; 71(4):2162–9.

84. Spring S, Schulze R, Overmann J, Schleifer KH. Identification and characterization of ecologically significant prokaryotes in the sediment of freshwater lakes: Molecular and cultivation studies. FEMS Microbiology Reviews. 2000. p. 573–90.

85. Hiorns WD, Hastings RC, Head IM, McCarthy AJ, Saunders JR, Pickup RW, et al. Amplification of 16S ribosomal RNA genes of autotrophic ammonia-oxidizing bacteria demonstrates the ubiquity of nitrosospiras in the environment. Microbiology. 1995; 141(11):2793–800.

86. Lindstrom ES and Leskinen E. Do neighboring lakes share common taxa of bacterioplankton? Comparison of 16S rDNA fingerprints and sequences from three geographic regions. Microb Ecol. 2002; 44(1):1–9.

87. Jian-hua Li Susumu Takii, Hidetake Hayashi KJP. Seasonal changes in ribosomal RNA of sulfate-reducing bacteria and sulfate reducing activity in a freshwater lake

sediment. FEMS Microbiol Ecol. 1999; 28:31–9.

88. Purdy KJ, Nedwell DB, Embley TM, Takii S. Use of 16S rRNA-targeted oligonucleotide probes to investigate the occurrence and selection of sulfate-reducing bacteria in response to nutrient addition to sediment slurry microcosms from a Japanese estuary. FEMS Microbiol Ecol. 1997; 24(3):221–34.

89. Wilms R, Sass H, Köpke B, Köster J, Cypionka H, Engelen B. Specific bacterial, archaeal, and eukaryotic communities in tidal-flat sediments along a vertical profile of several meters. Appl Environ Microbiol. 2006; 72(4):2756–64.

90. Winter C, Hein T, Kavka G, Mach RL, Farnleitner AH. Longitudinal changes in the bacterial community composition of the Danube River: A whole-river approach. Appl Environ Microbiol. 2007; 73(2):421–31.

91. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Applied and Environmental Microbiology. 2006. p. 1719–28.

92. Alexander M. Introduction to soil microbiology (2nd ed.). New York. John Wiley and Sons. 1977.

93. Weber S, Stubner S, Conrad R. Bacterial populations colonizing and degrading rice straw in anoxic paddy soil. Appl Environ Microbiol. 2001; 67(3):1318–27.

94. Felske A, Wolterink A, Van Lis R, De Vos WM, Akkermans ADL. Searching for predominant soil bacteria: 16S rDNA cloning versus strain cultivation. FEMS Microbiol Ecol. 1999; 30(2):137–45.

95. Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, et al. Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. Mol Microbiol. 2003; 50(2):563–75.

96. Whitman WB. Modern Soil Microbiology, second ed. Agric Syst. 2009;100(1-3): 95-96.

97. Liu W-T, Jansson K.J, Environmental Molecular Microbiology. UK. Caister Academic Press. Page 119-120.

98. Pisa G, Magnani GS, Weber H, Souza EM, Faoro H, Monteiro RA, et al. Diversity of 16S rRNA genes from bacteria of sugarcane rhizosphere soil. Brazilian J Med Biol Res. 2011; 44(12):1215–21.

99. Abou-Shanab RAI, van Berkum P, Angle JS, Delorme TA, Chaney RL, Ghozlan HA, et al. Characterization of Ni-resistant bacteria in the rhizosphere of the hyperaccumulator Alyssum murale by 16S rRNA gene sequence analysis. World J Microbiol Biotechnol. 2010; 26(1):101–8.

100. Araujo JF, de Castro AP, Costa MM, Togawa RC, Júnior GJ, Quirino BF, Bustamante MM, Williamson L, Handelsman J, Krüger RH. Characterization of

Soil Bacterial Assemblies in Brazilian Savanna-Like Vegetation Reveals Acidobacteria Dominance. Microb Ecol. 2012; 64(3):760–70.

101. Pereira RM, Silveira ÉL Da, Scaquitto DC, Aparecida E, Val-Moraes SP, Wickert E, et al. Molecular characterization of bacterial populations of different soils. Brazilian J Microbiol. 2006; 37(4):439–47.

102. C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM. Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J. 2009; 3(9):1004–11.

103. Fykse, E.M., Tjärnhage, T., Humppi, T. et al. Identification of airborne bacteria by 16S rDNA sequencing, MALDI-TOF MS and the MIDI microbial identification system. Aerobiologia (Bologna). 2015; 31(3):271–81.

104. Fahlgren C, Hagstro¨m A, Nilsson D, Zweifel U L. Annual variations in the diversity, viability, and origin of airborne bacteria. Appl Environ Microbiol. 2010; 76(9):3015–25.

105. Miletto M, Lindow SE. Relative and contextual contribution of different sources to the composition and abundance of indoor air bacteria in residences. Microbiome. 2015; 3(1):61.

106. Mancinelli RL, Shulls WA. Airborne bacteria in an urban environment. Appl Environ Microbiol. 1978; 35(6):1095–101.

107. Fang Z, Ouyang Z, Zheng H, Wang X, Hu L. Culturable airborne bacteria in outdoor environments in Beijing, China. Microb Ecol. 2007; 54(3):487–96.

108. Dybwad M, Granum PE, Bruheim PP, Blatnya JM. Characterization of airborne bacteria at an underground subway station. Appl Environ Microbiol. 2012; 78(6):1917–29.

109. Dybwad M, Skogan G, Blatnya JM. Temporal variability of the bioaerosol background at a subway station: Concentration level, size distribution, and diversity of airborne bacteria. Appl Environ Microbiol. 2014; 80(1):257–70.

110. Fykse EM, Langseth B, Olsen JS, Skogan G, Blatny JM. Detection of bioterror agents in air samples using real-time PCR. J Appl Microbiol. 2008; 105(2):351–8.

111. Peleg AY, Seifert H, Paterson DL. Acinetobacter baumannii: Emergence of a successful pathogen. Clinical Microbiology Reviews. 2008. p. 538–82.

112. Swann J, Richards S E, Shen Q, Holmes E, Marchesi J R & Tuohy K. Culture-Independent Analysis of the Human Gut Microbiota and their Activities, in Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats (ed F. J. de Bruijn). New Jersey. John Wiley & Sons. 2011. chapter 21; page 208.

113. Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoSONE. 2008; 3(7):e2836.

114. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, et al. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A. 2006; 103(3):732–7.

115. Reunamo A, Riemann L, Leskinen P, Jørgensen KS. Dominant petroleum hydrocarbon-degrading bacteria in the Archipelago Sea in South-West Finland (Baltic Sea) belong to different taxonomic groups than hydrocarbon degraders in the oceans. Mar Pollut Bul. Elsevier Ltd; 2013; 72(1):174–80.

116. Okoh AI. Biodegradation of Bonny light crude oil in soil microcosm by some bacterial strains isolated from crude oil flow stations saver pits in Nigeria. African J Biotechnol. 2003; 2(5):104–8.

117. Yrjälä K, Keskinen AK, Åkerman ML, Fortelius C, Sipilä TP. The rhizosphere and PAH amendment mediate impacts on functional and structural bacterial diversity in sandy peat soil. Environ Pollut. 2010; 158(5):1680–8.

118. Leahy JG, Colwell RR. Microbial degradation of hydrocarbons in the environment. Microbiol Rev. 1990; 54(3):305–15.

119. Chikere CB, Okpokwasili GC, Chikere BO. Bacterial diversity in a tropical crude oil-polluted soil undergoing bioremediation. J Biotechnol. 2009; 8(11):2535–40.

120. Hamamura N, Olson SH, Ward DM, Inskeep WP. Microbial population dynamics associated with crude-oil biodegradation in diverse soils. Appl Environ Microbiol. 2006; 72(9):6316–24.

121. Lin TC, Pan PT, Cheng SS. Ex situ bioremediation of oil-contaminated soil. J Hazard Mater. 2010; 176(1-3):27–34.

122. Alvarez VM, Marques JM, Korenblum E, Seldin L. Comparative Bioremediation of Crude Oil-Amended Tropical Soil Microcosms by Natural Attenuation, Bioaugmentation, or Bioenrichment. Appl Environ Soil Sci. 2011; 2011:1–10.

123. Baek, Kyung-Hwa, Byung-Dae Yoon, Byung-Hyuk Kim, Dae-Hyun Cho, In-Sook Lee, et al. Monitoring of microbial diversity and activity during bioremediation of crude oil-contaminated soil with different treatments. J Microbiol Biotechnol. 2007; 17(1):67–73.

124. Johnsen AR, Wick LY, Harms H. Principles of microbial PAH-degradation in soil. Environ Pollut. 2005; 133(1):71–84.

125. Head IM, Jones DM, Röling WFM. Marine microorganisms make a meal of oil. Nat Rev Microbiol. 2006; 4(3):173–82.

126. Wang W, Wang L, Shao Z. Diversity and abundance of oil-degrading bacteria and alkane hydroxylase (alkB) genes in the subtropical seawater of Xiamen Island. Microb Ecol. 2010; 60(2):429–39.

127. Al-Awadhi H, Al-Mailem D, Dashti N, Khanafer M, Radwan S. Indigenous hydrocarbon-utilizing bacterioflora in oil-polluted habitats in Kuwait, two decades after the greatest man-made oil spill. Arch Microbiol. 2012; 194(8):689–705.

128. Bælum J, Borglin S, Chakraborty R, Fortney JL, Lamendella R, Mason OU, et al. Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. Environ Microbiol. 2012; 14(9):2405–16.

129. Hazen TC, Dubinsky E a, DeSantis TZ, Andersen GL, Piceno YM, Singh N, et al. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science. 2010; 330(6001):204–8.

130. Abed RMM. Interaction between cyanobacteria and aerobic heterotrophic bacteria in the degradation of hydrocarbons. Int Biodeterior Biodegrad. 2010; 64(1):58–64.

131. Harayama S, Kasai Y, Hara A. Microbial communities in oil-contaminated seawater. Curr Opin Biotechnol. 2004; 15(3):205–14.

132. Ibraheem IBM. Biodegradability of hydrocarbons by cyanobacteria. J Phycol. 2010; 46(4):818–24.

133. Tang X, He LY, Tao XQ, Dang Z, Guo CL, Lu GN, et al. Construction of an artificial microalgal-bacterial consortium that efficiently degrades crude oil. J Hazard Mater. 2010; 181:1158–62.

134. McKew B a., Coulon F, Osborn a. M, Timmis KN, McGenity TJ. Determining the identity and roles of oil-metabolizing marine bacteria from the Thames estuary, UK. Environ Microbiol. 2007; 9(1):165–76.

135. Badin AL, Mustafa T, Bertrand C, Monier A, Delolme C, Geremia R a., et al. Microbial communities of urban stormwater sediments: The phylogenetic structure of bacterial communities varies with porosity. FEMS Microbiol Ecol. 2012; 81(2):324–38.

136. Vishnivetskaya T a., Mosher JJ, Palumbo A V., Yang ZK, Podar M, Brown SD, et al. Mercury and other heavy metals influence bacterial community structure in contaminated Tennessee streams. Appl Environ Microbiol. 2011; 77(1):302–11.

137. Zhu J, Zhang J, Li Q, Han T, Xie J, Hu Y, et al. Phylogenetic analysis of bacterial community composition in sediment contaminated with multiple heavy metals from the Xiangjiang River in China. Mar Pollut Bull. Elsevier Ltd; 2013; 70(1-2):134–9.

138. Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol. 2013; 79(6):1897–905.

139. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell, Fourth Edition. New York. Garland Science 2002.

140. Doolittle WF. Bacterial evolution. Can J Microbiol. 1988; 34(4):547–51.

141. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004; 17(4):840–862

142. Rajendhran J, Gunasekaran P. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiol Res. Elsevier; 2011; 166(2):99–110.

143. Woese CRCR, Magrum LJLJ, Gupta R, Siegel RB, Stahl D a., Kop J, et al. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Res. 1980; 8(10):2275–93.

144. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014; 12(9):635–45.

145. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 2007; 69(2):330–9.

146. Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: Evaluation of Effective Study Designs. PLoS One. 2013; 8(1):18–23.

147. Li H, Zhang Y, Li DS, Xu H, Chen GX, Zhang CG. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. Soil Biol Biochem. 2009; 41(5):954–68.

148. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol. 2005; 3(9):733–9.

149. Schloss PD. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput Biol. 2010; 6(7):19.

150. Weisburg WG, Barns SM, Pelletier D., Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991; 173(2):697–703.

151. Frank J a., Reich CI, Sharma S, Weisbaum JS, Wilson B a., Olsen GJ. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol. 2008; 74(8):2461–70.

152. Kuczynski J, Lauber CL, Walters W a., Parfrey LW, Clemente JC, Gevers D, et al. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet. 2011; 13(1):47–58.

153. Armougom F, Raoult D. Exploring microbial diversity using 16S rRNA high-throughput methods. J Comput Sci Syst Biol. 2009; 2(1):74–92.

154. Soergel D a W, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 2012; 6(7):1440–4.

155. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010; 38(22).

156. Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, et al. Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. DNA Res. 2014; 21(2):217–27.

157. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013; 41(1):1–11.

158. Nossa CW, Oberdorf WE, Yang L, Aas J a., Paster BJ, de Santis TZ, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol. 2010; 16(33):4135–44.

159. Cai L, Ye L, Tong AHY, Lok S, Zhang T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One. 2013; 8(1):1–11.

160. Stahl DA, Lane DJ, Olsen GJ, Pace NR. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. Science. 1984; 224(4647):409–11.

161. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A. 1985; 82(20):6955–9.

162. Pace NR, Stahl DA, Lane DJ OG. The analysis of natural microbial-populations by ribosomal-RNA sequences. Adv Microb Ecol. 1986; 9:1-55.

163. Schmidt TM, DeLong EF, Pace NR. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol. 1991; 173(14):4371–8.

164. Devereux R, Mundfrom GW. A phylogenetic tree of 16S rRNA sequences from sulfate-reducing bacteria in a sandy marine sediment. Appl Environ Microbiol. 1994; 60(9):3437–9.

165. Zhou J, Davey ME, Figueras JB, Rivkina E, Gilichinsky D, Tiedje JM. Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. Microbiology. 1997; 143 (12):3913–9.

166. Weller R, Bateson MM, Heimbuch BK, Kopczynski ED, Ward DM. Uncultivated cyanobacteria, Chloroflexus-like inhabitants, and spirochete-like inhabitants of a hot spring microbial mat. Appl Environ Microbiol. 1992; 58(12):3964–9.

167. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995; 59(1):143–69.

168. Hugenholtz P, Tyson GW. Metagenomics. Nature. 2008; 455:481–3.

169. Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol Rev. 1994; 15(2-3):155–73.

170. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen J a, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004; 304(5667):66–74.

171. Lo A, Salgado H, Martı I, Solano H, Collado-vides J. Bacterial molecular networks. Bact Mol Networks Methods Protoc. 2012; 804:179–95.

172. Handelsman J. Metagenomics. Application of genomics to uncultured Microorganisms. Microbiol Mol Biol Rev. 2005; 69(1):195–195.

173. Tringe SG, von Mering C, Kobayashi A, Salamov A, Chen K, Chang HW. Comparative metagenomics of microbial communities. Science. 2005; 308(5721):554–7.

174. Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V, Batra N. Metagenomics: Concept, methodology, ecological inference and recent advances. Biotechnol J. 2009; 4(4):480-94.

175. Xu J. Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. Mol Ecol. 2006; 15(7):1713–31.

176. Simon C, Daniel R. Metagenomic analyses: Past and future trends. Appl Environ Microbiol. 2011; 77(4):1153–61.

177. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol. 2004; 186(9):2629–35.

178. Gich FB, Amer E, Figueras JB, Abella C a, Balaguer MD, Poch M. Assessment of microbial community structure changes by amplified ribosomal DNA restriction analysis (ARDRA). Int Microbiol. 2000; 3(2):103–6.

179. Lagacé L, Pitre M, Jacqeus M, Roy D. Identification of the bacterial community of maple sap by using amplified ribosomal DNA (rDNA) restriction analysis and rDNA sequencing. Appl Environ Microbiol. 2004; 70(4):2052–60.

180. Wu XY, Walker MJ, Hornitzky M, Chin J. Development of a group-specific PCR

combined with ARDRA for the identification of Bacillus species of environmental significance. J Microbiol Methods. 2006; 64(1):107–19.

181. Sievert SM, Kuever J, Muyzer G. Identification of 16s ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos island (Greece). Appl Environ Microbiol. 2000; 66(7):3102–9.

182. Brümmer IHM, Felske A, Wagner-Döbler I. Diversity and seasonal variability of b-proteobacteria in biofilms of polluted rivers: Analysis by temperature gradient gel electrophoresis and cloning. Appl Environ Microbiol. 2003; 69(8):4463–73.

183. Marshall SM, Melito PL, Woodward DL, Johnson WM, Rodgers FG, Mulvey MR. Rapid identification of Campylobacter, Arcobacter, and Helicobacter isolates by PCR-restriction fragment length polymorphism analysis of the 16S rRNA gene. J Clin Microbiol. 1999; 37(12):4158–60.

184. Dahllöf I. Molecular community analysis of microbial diversity. Curr Opin Biotechnol. 2002; 13(3):213–7.

185. Spiegelman D, Whissell G, Greer CW. A survey of the methods for the characterization of microbial consortia and communities. Can J Microbiol. 2005; 51(5):355–86.

186. The Science Creative Quarterly. Denaturing Gradient Gel Electrophoresis (DGGE): An Overview.
Retrieved from URL: http://www.scq.ubc.ca/denaturing-gradient-gel-electrophoresis-dgge-an-overview/

187. Denaturing Gradient Gel Electrophoresis (DGGE). Laboratory for Microbial Ecology, Department of Earth, Ecological and Environmental Sciences, University of Toledo.
Retrieved from URL:
http://www.eeescience.utoledo.edu/Faculty/Sigler/Von_Sigler/LEPR_Protocols_files/DGGE.pdf

188. Pontes DS, Lima-Bittencourt CI, Chartone-Souza E, Amaral Nascimento AM. Molecular approaches: Advantages and artifacts in assessing bacterial diversity. J Ind Microbiol Biotechnol. 2007; 34(7):463–73.

189. Williamson SJ, Yooseph S. From bacterial to microbial ecosystems (Metagenomics). Methods Mol Biol. 2012; 804:35–55.

190. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010; 11(1):31–46.

191. Nyrén P. The history of pyrosequencing. Methods Mol Biol. 2007; 373(3):1–14.

192. Sanger, F. and A.R. Coulson, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol, 1975; 94(3): 441-8.

193. Sanger, F., S. Nicklen, and A.R. Coulson, DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 1977; 74(12): 5463-7.

194. Ronaghi, M., M. Uhlen, and P. Nyren, A sequencing method based on real-time pyrophosphate. Science, 1998; 281(5375): 363, 365.

195. Ronaghi, M., et al., Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem, 1996; 242(1): 84-9.

196. TOTTY, M., A better Idea. The Wall Street Journal. October, 2005. Retrieved from URL: http://www.wsj.com/articles/SB112975757605373586

197. Jones WJ. High-throughput sequencing and metagenomics. Estuaries and Coasts. 2010;33(4):944–52.

198. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008; 9:387–402.

199. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben L a, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437(7057):376–80. 0

200. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol. 2008; 26(10):1117–24.

201. Strausberg RL, Levy S, Rogers Y-H. Emerging DNA sequencing technologies for human genomic medicine. Drug Discov Today. 2008; 13(13-14):569–77.

202. Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J, et al. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. Environ Microbiol. 2012; 14(9):2457–66.

203. Robertson GP, Klingensmith KM, Klug MJ, Paul E a, Crum JR, Ellis BG. Soil resources, microbial activity, and primary production across an agricultural ecosystem. Ecological Applications. 1997. p. 158–70.

204. Barreto DP, Conrad R, Klose M, Claus P, Enrich-Prast A. Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. PLoS One. 2014; 9(10):e110128.

205. Tebbe CC, Vahjen W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl Environ Microbiol. 1993; 59(8):2657–65.

206. LaMontagne MG, Michel FC, Holden P a., Reddy C a. Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. J Microbiol Methods. 2002; 49(3):255–64.

207. Zipper H, Buta C, Lämmle K, Brunner H, Bernhagen J, Vitzthum F. Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. Nucleic Acids Res.

2003; 31(7):e39.

208. Lakay FM, Botha a., Prior B a. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. J Appl Microbiol. 2007; 102(1):265–73.

209. Takada-Hoshino Y, Matsumoto N. An improved DNA extraction method using skim milk from soils that strongly adsorb DNA. Microbes Environ. 2004; 19(1):13–9.

210. Kim W, Simkin M, Ph D. Application note 46 DNA sample preparation: the effect of sample preparation on humic acid removal from peat samples : A Comparative Study . Page 2–3.

211. Morales SE, Holben WE. Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. Appl Environ Microbiol. 2009; 75(9):2677–83.

212. Suzuki MT, Giovannoni SJ. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microbiol. 1996; 62(2):625–30.

213. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol. 1998; 64(10):3724–30.

214. Thompson JR, Marcelino L a, Polz MF. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by "reconditioning PCR". Nucleic Acids Res. 2002; 30(9):2083–8.

215. Kopczynski ED, Bateson MM, Ward DM. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl Environ Microbiol. 1994; 60(2):746–8.

216. Liesack W, Weyland H, Stackebrandt E. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microb Ecol. 1991; 21(1):191–8.

217. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward D V., Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 2011; 21:494–504.

218. Ahn JH, Kim BY, Song J, Weon HY. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. J Microbiol. 2012; 50(6):1071–4.

219. Hugenholtz P, Huber T. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. Int J Syst Evol Microbiol. 2003; 53:289–93.

220. Qiu X, Wu L, Huang H, McDonel PE, Palumbo A V, Tiedje JM, et al. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-

based cloning. Appl Environ Microbiol. 2001; 67(2):880–7.

221. Brandariz-Fontes C, Camacho-Sanchez M, Vilà C, Vega-Pla JL, Rico C, Leonard JA. Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. Sci Rep. 2015; 5:8056.

222. Stackebrandt E, Pukall R, Ulrichs G, Rheims H. Analysis of 16S rDNA clone libraries: part of the big picture. Proc 8th Int Symp Microb Ecol Microb Biosyst new Front Atl Canada Soc Microb Ecol Halifax, Nov Scotia, Canada. 1999; 1–9.

223. Klappenbach J, Saxman PR, Cole JR, Schmidt TM. rrndb: the Ribosomal RNA Operon Copy Number Database. Nucleic Acids Res. 2001; 29(1):181–4.

224. Klappenbach J, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol. 2000; 66(4):1328–33.

225. Asai T, Condon C, Voulgaris J, Zaporojets D, Shen B, Al-Omar M, et al. Construction and initial characterization of Escherichia coli strains with few or no intact chromosomal rRNA operons. J Bacteriol. 1999; 181(12):3803–9.

226. Fegatella F, Lim J, Kjelleberg S, Cavicchioli R. Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium Sphingomonas sp. strain RB2256. Appl Environ Microbiol. 1998; 64(11):4433–8.

227. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz E a., et al. Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl Environ Microbiol. 2010; 76(12):3886–97.

228. Rainey F a., Ward-Rainey NL, Janssen PH, Hippe H, Stackebrandt E. Clostridium paradoxum DSM 7308(T) contains multiple 16S rRNA genes with heterogeneous intervening sequences. Microbiology. 1996; 142(8):2087–95.

229. Rastogi R, Wu M, Dasgupta I, Fox GE. Visualization of ribosomal RNA operon copy number distribution. BMC Microbiol. 2009; 9:208.

230. Huse SM, Welch DM, Morrison HG, Sogin ML. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol. 2010; 12(7):1889–98.

231. Beuf K De, Schrijver J De, Thas O, Criekinge W Van, Irizarry R a, Clement L. Improved base-calling and quality scores for 454 sequencing based on a Hurdle Poisson model. BMC Bioinformatics. 2012; 13:303.

232. Balzer S, Malde K, Lanzén A, Sharma A, Jonassen I. Corrigendum: Characteristics of 454 pyrosequencing data-enabling realistic simulation with flowsim. Bioinformatics. 2011; 27(15):2171.

233. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics. 2011; 12(1):245.

234. Ward D V., Gevers D, Giannoukos G, Earl AM, Methé B a., Sodergren E, et al. Evaluation of 16s rDNA-based community profiling for human microbiome research. PLoS One. 2012; 7(6):e39315.

235. Anh MT, Do Hong LC, Nguyen NV, Tu Thi CL, Minh TL, Becker-Van Slooten K. Micropollutants in the sediment of the Saigon-Dongnai river: Situation and ecological risks. Chimia. 2003; 57(9):537–41.

236. Minh Ngoc N , My Binh P, Tuyet Nga N*, Phi Nga L. Ô nhiễm PAHs và khả năng phân huỷ các hợp chất hydrocarbon thơm bởi quần thể vi khuẩn bùn sông Sài Gòn. Jour of Scien & Tech. 2014; 52: 262-73.

237. Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. Nat Protoc. 2008; 3(2):267–78.

238. Zheng Z, Advani A, Melefors O, Glavas S, Nordstrom H, Ye W, et al. Titration-free massively parallel pyrosequencing using trace amounts of starting material. Nucleic Acids Res. 2010; 38(13):137.

239. An S, Couteau C, Luo F, Neveu J, DuBow MS. Bacterial diversity of surface sand samples from the Gobi and Taklamaken Deserts. Microb Ecol. 2013; 66(4):850–60.

240. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009; 75(23):7537–41.

241. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16s rRNA-based studies. PLoS One. 2011; 6(12).

242. Martin M. Sequencing reads. EMBnet.journal. 2011; 17.1:10–12.

243. Smeds L KA. ConDeTri--a content dependent read trimmer for Illumina data. PLoS One. 2011; 6(10):e263.

244. Wright ES, Yilmaz LS, Noguera DR. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol. 2012; 78(3):717–25.

245. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011; 27(16):2194–200.

246. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. Nature Publishing Group; 2012; 6(3):610–8.

247. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive analysis and graphics of microbiome census data. PLoS One. 2013; 8(4).

248. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28(23):3150–2.

249. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013; 41: D590-6.

250. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014; 42:D643–8.

251. Chessel D, Dufour AB, Thioulouse J. The ade4 package - I : One-table methods. R News. 2004; 4(1):5–10.

252. Method 1681: Fecal Coliforms in Sewage Sludge (Biosolids) by MultipleTube Fermentation using A-1 medium July 2006.
Retrieved from URL:
http://water.epa.gov/scitech/methods/cwa/bioindicators/upload/2008_11_25_methods_method_biological_1681_1.pdf

253. Feng PCS, Hartman PA. Fluorogenic assays for immediate confirmation of Escherichia coli. Appl Environ Microbiol. 1982;43(6):1320–9.

254. Sutton S. The Most Probable Number Method and its uses in enumeration, qualification, and validation. J Valid Technol. 2010;16(3):35–8.

255. Method SMEWW 5530-Phenols:2005.
Retrieved from URL:
https://www.standardmethods.org/store/ProductView.cfm?ProductID=42

256. EPA Method 3050B: Acid Digestion of Sediments, Sludges, and Soils.
Retrieved from URL: https://www.epa.gov/sites/production/files/2015-06/documents/epa-3050b.pdf

257. Method SMEWW 3120 B – IC .
Retrieved from URL:
https://www.standardmethods.org/store/ProductView.cfm?ProductID=210

258. IUPAC code is available at http://www.bioinformatics.org/sms/iupac.html

259. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014; 42(D1).

260. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010; 11(1):31–46.

261. 454 Sequencing System Software Manual Version 2.9.
Retrieved from URL: http://gyra.ualg.pt/misc/USM-00058.09_454SeqSys_SWManual-v2.9_PartB.pdf

262. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Publ Gr. 2010; 7(5):335–6.

263. Blankenberg D, Coraor N, Von Kuster G, Taylor J, Nekrutenko A. Integrating diverse databases into unified analysis framework: A Galaxy approach. Database. 2011; 2011.

264. Avramidis P, Nikolaou K, Bekiari V. Total organic carbon and total nitrogen in sediments and Soils: A comparison of the wet oxidation – titration method with the combustion-infrared method. Agric Agric Sci Procedia. 2015; 4:425–30.

265. Retrieved from URL: https://www.ontario.ca/page/soil-ground-water-and-sediment-standards-use-under-part-xv1-environmental-protection-act

266. Canadian Sediment Quality Guidelines for the Protection of Aquatic Life – Update 2002.
Retrieved from URL: https://www.pla.co.uk/Environment/Canadian-Sediment-Quality-Guidelines-for-the-Protection-of-Aquatic-Life

267. Browning SR, Browning BL. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. Am J Hum Genet. 2015; 97(3):404–18.

268. Hill MO. Diversity and Evenness: A unifying notation and its consequences. Ecology. 1973; 54(2):427–32.

269. Minh NH, Minh TB, Iwata H, Kajiwara N, Kunisue T, Takahashi S, et al. Persistent organic pollutants in sediments from Sai Gon-Dong Nai River basin, Vietnam: Levels and temporal trends. Arch Environ Contam Toxicol. 2007; 52(4):458–65.

270. Lao W, Tsukada D, Greenstein DJ, Bay SM, Maruya KA. Analysis, occurrence, and toxic potential of pyrethroids, and fipronil in sediments from an urban estuary. Environ Toxicol Chem. 2010; 29(4):843–51.

271. Li X, Liu L, Wang Y, Luo G, Chen X, Yang X, et al. Integrated assessment of heavy metal contamination in sediments from a coastal industrial basin, NE China. PLoS One. 2012; 7(6).

272. Barakat a, Baghdadi M El. Assessment of Heavy metal in surface sediments of Day River at Beni-Mellal Region, Morocco. Res Jour of Environ and Earth Scien. 2012; 4(8):797–806.

273. Thongra-ar W, , Chaluay Musika WW and AM. Heavy metals contamination in sediments along the Eastern Coast of the Gulf of Thailand. Environ asia. 2008; 1:37–45.

274. Kafilzadeh F. Distribution and sources of polycyclic aromatic hydrocarbons in water and sediments of the Soltan Abad River, Iran. Egypt J Aquat Res. 2015; 41(3):227–31.

275. Vane CH, Kim AW, Beriro DJ, Cave MR, Knights K, Moss-Hayes V, et al. Polycyclic aromatic hydrocarbons (PAH) and polychlorinated biphenyls (PCB) in urban soils of Greater London, UK. Appl Geochemistry. 2014; 51:303–14.

276. Canadian Council of Ministers of the Environment. Canadian Soil Quality Guidelines for carcinogenic and other polycyclic aromatic hydrocarbons (Environmental and Human Health Effects); 2008.

277. Hafner C, Gartiser S, Garcia-Käufer M, Schiwy S, Hercher C, Meyer W, et al. Investigations on sediment toxicity of German rivers applying a standardized bioassay battery. Environ Sci Pollut Res. 2015; 22(21):16358–70.

278. Ibekwe AM, Ma J, Murinda SE. Bacterial community composition and structure in an Urban River impacted by different pollutant sources. Sci Total Environ. 2016; 566-567:1176–85.

279. Xu HJ, Li S, Su JQ, Nie S, Gibson V, Li H, et al. Does urbanization shape bacterial community composition in urban park soils? A case study in 16 representative Chinese cities based on the pyrosequencing method. FEMS Microbiol Ecol. 2014; 87(1):182–92.

280. Jordaan K, Bezuidenhout CC. Bacterial community composition of an urban river in the North West Province, South Africa, in relation to physico-chemical water quality. Environ Sci Pollut Res Int. 2016; 23(6):5868–80.

281. Zhao D, Huang R, Zeng J, Yan W, Wang J, Ma T, et al. Diversity analysis of bacterial community compositions in sediments of urban lakes by terminal restriction fragment length polymorphism (T-RFLP). World J Microbiol Biotechnol. 2012; 28(11):3159–70.

282. Ligi T, Oopkaup K, Truu M, Preem J-K, Nolvak H, Mitsch W, et al. Characterization of bacterial communities in soil and sediment of a created riverine wetland complex using high-throughput 16s rRNA amplicon sequencing. Ecol Eng. 2013.

283. Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, et al. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 2007.

284. Garrity GM, Holt JG, Spieck E, Bock E, Johnson DB, Spring S, et al. Phylum BVIII. Nitrospirae phy. nov. Bergey's Manual of Systematic Bacteriology. 2001. p. 457–60.

285. Hanada S. The phylum chloroflexi, the family chloroflexaceae, and the related phototrophic families oscillochloridaceae and roseiflexaceae. The Prokaryotes: Other Major Lineages of Bacteria and The Archaea. 2014. p. 515–32.

286. Zhao D, Huang R, Zeng J, Yan W, Wang J, Ma T, et al. Diversity analysis of bacterial community compositions in sediments of urban lakes by terminal restriction fragment length polymorphism (T-RFLP). World J Microbiol Biotechnol. 2012; 28(11):3159–70.

287. H. Z, S. D, W. S, G. W, Y. G, J. G, et al. Comparison of bacterial diversity of polluted and unpolluted sediment by brominated flame retardant. Wei Sheng Wu Xue Bao. 2011; 51(3):377–85.

288. Björnsson L, Hugenholtz P, Tyson GW, Blackall LL. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. Microbiology. 2002; 148(8):2309–18.

289. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome. 2013; 1(1):22.

290. Robinson G, Caldwell GS, Wade MJ, Free A, Jones CLW, Stead SM. Profiling bacterial communities associated with sediment-based aquaculture bioremediation systems under contrasting redox regimes. Sci Rep. 2016; 6(1):38850.

291. Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol. 2013; 79(6):1897–905.

292. Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, et al. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev. 2007;71(3):495–548.

293. Gołębiewski M, Deja-Sikora E, Cichosz M, Tretyn A, Wróbel B. 16S rDNA Pyrosequencing Analysis of Bacterial Community in Heavy Metals Polluted Soils. Microb Ecol. 2014; 67(3):635–47.

294. Altmann D, Stief P, Amann R, De Beer D, Schramm A. In situ distribution and activity of nitrifying bacteria in freshwater sediment. Env Microbiol [Internet]. 2003; 5(9):798–803.

295. Daims H, Nielsen JL, Nielsen PH, Schleifer KH, Wagner M. In situ characterization of Nitrospira-like nitrite-oxidizing bacteria active in wastewater treatment plants. Appl Environ Microbiol [Internet]. 2001; 67(11):5273–84.

296. Hovanec TA, Taylor LT, Blakis A, Delong EF. Nitrospira-like bacteria associated with nitrite oxidation in freshwater aquaria. Appl Environ Microbiol. 1998; 64(1):258–64.

297. Hayyan M, Sameh S a., Hayyan A, AlNashef IM. Utilizing of sodium nitrite as inhibitor for protection of carbon steel in salt solution. Int J Electrochem Sci. 2012; 7(8):6941–50.

298. Karim S, Mustafa CM, Assaduzzaman M, Islam M. Effect of nitrate ion on corrosion inhibition of mild steel in simulated cooling water. Chem Eng Res Bull. 2010; 14(2):87–91.

299. Wachter a. Sodium Nitrite as Corrosion Inhibitor for Water. Ind Eng Chem. 1945; 9–11.

300. Cébron A, Garnier J. Nitrobacter and Nitrospira genera as representatives of nitrite-oxidizing bacteria: detection, quantification and growth along the lower Seine River (France). Water Res. 2005; 39(20):4979–92.

301. Nakamura Y, Satoh H, Kindaichi T, Okabe S. Community structure, abundance, and in situ activity of nitrifying bacteria in river sediments as determined by the combined use of molecular techniques and microelectrodes. Environ Sci Technol. 2006; 40(5):1532–9.

302. Raimonet M, Vilmin L, Flipo N, Rocher V, Laverman AM. Modelling the fate of nitrite in an urbanized river using experimentally obtained nitrifier growth parameters. Water Res. 2015; 73:373–87.

303. Institute for Global Environmental Strategies. Water resources management in Ho Chi Minh City. Sustain Groundw Manag Asian Cities. 2007; 68–92.

304. EUZéBY JP. List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. Int J Syst Evol Microbiol. 1997; 47(2):590–2.

305. Hallberg KB, Hedrich S, Johnson DB. Acidiferrobacter thiooxydans, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron- and sulfur-oxidizer of the family Ectothiorhodospiraceae. Extremophiles. 2011;15(2):271–9.

306. Harrison AP. Genomic and physiological diversity amongst strains of *Thiobacillus ferrooxidans*, and genomic comparison with *Thiobacillus thiooxidans*. Arch Microbiol. 1982; 131(1):68–76.

307. Kojima H, Shinohara A, Fukui M. Sulfurifustis variabilis gen. nov., sp. nov., a sulfur oxidizer isolated from a lake, and proposal of Acidiferrobacteraceae fam. nov. and Acidiferrobacterales ord. nov. Int J Syst Evol Microbiol. 2015;65(10):3709–13.

308. Dyksma S, Bischof K, Fuchs BM, Hoffmann K, Meier D, Meyerdierks A, et al. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. ISME J. 2016;10(8):1939–53.

309. Nowak ME, Schwab VF, Lazar CS, Behrendt T, Kohlhepp B, Totsche KU, et al. Carbon isotopes of dissolved inorganic carbon reflect utilization of different carbon sources by microbial communities in two limestone aquifer assemblages. Hydrol Earth Syst Sci Discuss. 2016; 1–36.

310. Niu J, Deng J, Xiao Y, He Z, Zhang X, Van Nostrand JD, Liang Y, Deng Y, Liu X, Yin H. The shift of microbial communities and their roles in sulfur and iron cycling in a copper ore bioleaching system. Sci Rep. 2016 Oct 4; 6:34744.

311. Zhang J, Sun QL, Zeng ZG, Chen S, Sun L. Microbial diversity in the deep-sea sediments of Iheya North and Iheya Ridge, Okinawa Trough. Microbiol Res. 2015;177:43–52.

312. Fan M, Lin Y, Huo H, Liu Y, Zhao L, Wang E, Chen W, Wei G. Microbial communities in riparian soils of a settling pond for mine drainage treatment. Water Res. 2016 Jun 1; 96:198-207.

313. Ye J, Joseph SD, Ji M, Nielsen S, Mitchell DRG, Donne S, Horvat J, Wang J, Munroe P, Thomas T. Chemolithotrophic processes in the bacterial communities on the surface of mineral-enriched biochars. ISME J. 2017 May; 11(5):1087-1101.

314. McIlroy Simon J., Kirkegaard Rasmus H., Dueholm Morten S., Fernando Eustace, Karst M., Albertsen Mads, Nielsen Per H. Culture-independent analyses reveal novel Anaerolineaceae as abundant primary fermenters in anaerobic digesters Treating Waste Activated Sludge. Frontier in Microbiolgy, Front. Microbiol., 23 June 2017.

315. Zheng B, Wang L, Liu L. Bacterial community structure and its regulating factors in the intertidal sediment along the Liaodong Bay of Bohai Sea, China. Microbiol Res. 2014; 169(7-8):585–92.

316. Mendes LW, Tsai SM. Variations of bacterial community structure and composition in mangrove sediment at different depths in Southeastern Brazil. Diversity. 2014; 6(4):827–43.

317. Gao F, Li F, Tan J, Yan J, Sun H. Bacterial community composition in the gut content and ambient sediment of sea cucumber Apostichopus japonicus revealed by 16S rRNA gene pyrosequencing. PLoS One. 2014 Jun 26; 9(6):e100092.

318. Kim JS, Lee KC, Kim DS, Ko SH, Jung MY, Rhee SK, et al. Pyrosequencing analysis of a bacterial community associated with lava-formed soil from the Gotjawal forest in Jeju, Korea. Microbiologyopen. 2015; 4(2):301–12.

319. Jordaan, K. & Bezuidenhout, C.C. Bacterial community composition of an urban river in the North West Province, South Africa, in relation to physico-chemical water quality.Environ Sci Pollut Res; 2016 23: 5868.

320. García-Armisen T, İnceoğlu Ö, Ouattara NK, Anzil A, Verbanck MA, et al. Seasonal variations and resilience of bacterial communities in a sewage polluted urban river. PLOS ONE; 2014, 9(3): e92579.

321. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007; 35(9):3100–8.

322. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, et al. Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl Environ Microbiol. 2010; 76(12):3886–97.

323. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol. 2012; 8(10).

324. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol. 2012; 8(10).

325. Wang L, Liu L, Zheng B, Zhu Y, Wang X. Analysis of the bacterial community in the two typical intertidal sediments of Bohai Bay, China by pyrosequencing. Mar Pollut Bull. Elsevier Ltd; 2013; 72(1):181–7.

326. Kwon S, Moon E, Kim T-S, Hong S, Park H-D. Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. Microbes Environ. 2011; 26(2):149–55.

327. Cai L, Ye L, Tong AHY, Lok S, Zhang T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One. 2013; 8(1):1–11.

328. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One. 2012; 7(8):e43093.

329. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011; 108 :4516–22.

330. López-Lozano NE, Heidelberg KB, Nelson WC, García-Oliva F, Eguiarte LE, Souza V. Microbial secondary succession in soil microcosms of a desert oasis in the Cuatro Cienegas Basin, Mexico. PeerJ. 2013; 1:e47.

331. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013; 41(1):1–11.

332. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Welch DM, Relman DA, et al. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet [Internet]. 2008; 4(11):e1000255.

333. Bowen de León K, Ramsay BD, Fields MW. Quality-score refinement of SSU rRNA gene pyrosequencing differs across gene region for environmental samples. Microb Ecol. 2012; 64(2):499–508.

334. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16s rRNA-based studies. PLoS One. 2011; 6(12).

335. Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, de Santis TZ, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol. 2010; 16(33):4135–44.

336. Cai L, Ye L, Tong AHY, Lok S, Zhang T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One. 2013; 8(1).

337. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013; 41(1).

338. Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl Environ Microbiol. 2009; 75(16):5227–36.

339. Kumar PS, Brooker MR, Dowd SE, Camerlengo T. Target region selection is a critical determinant of community fingerprints generated by 16S Pyrosequencing. PLoS One. 2011; 6(6).

340. Hitch RK, Day HR. Unusual persistence of DDT in some Western USA soils. Bull Environ Contam Toxicol. 1992; 48(2):259-64.

341. Spencer W, Singh G, Taylor C, LeMert R, Cliath M, Farmer W. DDT persistence and volatility as affected by management practices after 23 years. J Environ Qual. 1996; 25:815–821.

342. Wedemeyer G. Dechlorination of 1,1,1-trichloro-2,2-bis(p-chlorophenyl)ethane by Aerobacter aerogenes. Appl Microbiol. 1967; 15(3):569-4.

343. Kelce WR, Stone CR, Laws SC, Gray LE, Kemppainen JA, Wilson EM. Persistent DDT metabolite p,p'-DDE is a potent androgen receptor antagonist. Nature. 1995; 375(6532):581-5.

344. Schmidt TSB, Matias Rodrigues JF, von Mering C. Ecological Consistency of SSU rRNA-Based Operational Taxonomic Units at a Global Scale. PLoS Comput Biol. 2014; 10(4).

345. Chao A. Nonparametric Estimation of the Number of Classes in a Population Author. Scanadinavian J Stat. 1984;11(4):265–70.

346. Abhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics. 2015; 31(17):2882–4.

347. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006; 72(7):5069–72.

348. Kunin V, Hugenholtz P. PyroTagger: A fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. Open Journal. 2010.

349. Giongo A, Crabb DB, Davis-Richardson AG, Chauliac D, Mobberley JM, Gano KA, et al. PANGEA: pipeline for analysis of next generation amplicons. ISME J. 2010; 4(7):852–61.

350. Angiuoli SV, Matalka M, Gussman A, Galens K, Vangala M, et al. CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. BMC Bioinformatics. 2011; 12:356.

351. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27(6):863–4.

352. Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren A, et al. VAMPS: a website for visualization and analysis of microbial population structures. BMC Bioinformatics. 2014; 15(1):41.

353. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12(1):444.

354. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007; 17(3):377–86.

355. Bowen de León K, Ramsay BD, Fields MW. Quality-score refinement of SSU rRNA gene pyrosequencing differs across gene region for environmental samples. Microb Ecol. 2012; 64(2):499–508.

356. Dodt M, Roehr J, Ahmed R, Dieterich C. FLEXBAR—Flexible barcode and adapter processing for Next-Generation Sequencing platforms. Biology (Basel). 2012; 1(3):895–905.

357. Plotting PCA (Principal Component Analysis).
Retrieved from:
https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html

358. Dong C Di, Chen CF, Chen CW. Determination of polycyclic aromatic hydrocarbons in industrial harbor sediments by GC-MS. Int J Environ Res Public Health. 2012; 9(6):2175–88.

Résumé

La pollution de l'eau menace la santé publique et a un impact négatif sur le système écologique des rivières (**33**). Le système des rivières Saigon et Dong Nai est la plus importante source d'eau pour les douze villes et provinces du sud du Vietnam. Il est aujourd'hui gravement pollué par les activités humaines, agricoles, industrielles et domestiques, constituant une menace pour la vie de millions de personnes. Les rivières traversent 10 provinces et Hô Chi Minh Ville (HCMV) où 103 zones industrielles se situent avec plus de 19 millions d'habitants (figure A). L'eau, dans certaines sections en aval du système des rivières Saigon et Dong Nai a dépassé les niveaux de danger, selon le Département de contrôle de la pollution (PCD) du Ministère des ressources naturelles et de l'environnement (**14**). Le ministère vietnamien des Ressources naturelles et de l'environnement a rapporté que les rivières ont reçu environ 1,54 milliards de litres d'eaux usées provenant de 70 parcs industriels par jour, dont 35% de déchets médicaux non traités, et que des tests effectués depuis 2006 ont montré des niveaux élevés de pollution en particulier de substances toxiques organiques (**14**). Pour cette raison, la qualité de l'approvisionnement en eau est une préoccupation pour le gouvernement. En 2009, le vice-président de la compagnie WASACO, chargée de l'approvisionnement en eau de HCMV, a averti qu'avec une telle augmentation de la pollution, l'approvisionnement en eau par la compagnie WASACO sera bientôt impossible. De plus, la vie aquatique ne pourra plus survivre en raison des niveaux élevés de pollution (**14**).

Les micro-organismes sont essentiels au recyclage des nutriments dans les écosystèmes en raison de leur capacité à décomposer les composés organiques. Par conséquent, ils agissent comme une partie essentielle pour décider de la qualité de l'eau dans la rivière. De plus, les microbes s'avèrent être le premier groupe d'organismes vivants affectés par les changements environnementaux. Jusqu'à présent, il n'y a pas de données sur la diversité microbienne dans le système fluvial des rivières Saigon et Dong Nai, en particulier dans les sédiments, où se trouve généralement la plus grande partie de la biomasse microbienne d'une rivière. De plus, les sédiments fluviaux se sont révélés être un réservoir possible de pathogènes pouvant affecter la santé publique.

Dans le cadre de ce projet de thèse, nous nous sommes aux différents composés chimiques polluant la rivière ainsi qu'à la diversité microbienne dans les rivières Saigon et Dong Nai. Dans un premier temps, nous avons déterminé la diversité

microbienne. La méthode la plus couramment utilisée pour révéler toute la diversité microbienne dans un environnement donné est d'examiner les séquences de l'ADNr 16S, possédées par tous les procaryotes (figure B). Dans le cadre de ce projet de thèse, 13 sites, situées le long des rivières Saigon et Dong Nai ont été choisi afin de déterminer la diversité microbienne le long des rivières.* Afin de caractériser les populations microbiennes présentes sur les 13 sites choisis, un total de 42 échantillons de sédiments ont été prélevés (figure 2.4) parmi les différents sites. Puis, l'ADN total de chaque échantillon environnemental a été extrait et amplifié dans les régions V3 à V1 de l'ADNr 16S. L'ADN amplifié a ensuite été séquencé par la méthode de pyroséquençage. Les résultats obtenus grâce à la méthode de pyroséquençage ont été analysés par des approches bio-informatiques afin de déterminer la diversité microbienne des différents échantillons.

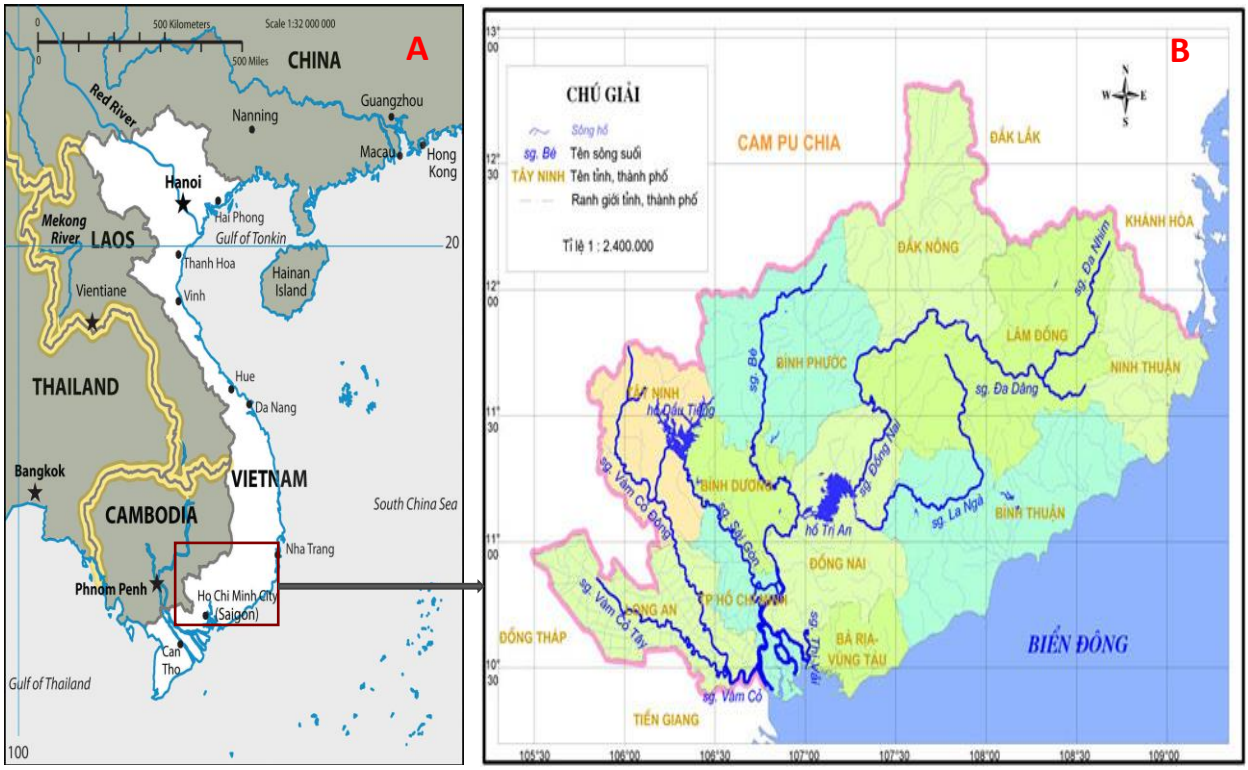


Figure A. Le système de rivières Saigon et Dong Nai (SG-DN).

A) Position du système fluvial SG-DN au Vietnam (**6, 7**).

B) Le système fluvial SG-DN qui traverse HCMV et 10 provinces du Vietnam (**8**).

*Les 13 sites étudiés correspond à des régions très polluées des rivières Saigon et Dong Nai. Parmi les 13 sites, 5 sont situés le long de la rivière Saigon de l'amont vers l'aval, 5 sont situés le long de la rivière Dong Nai de l'amont à l'aval correspond à l'intersection

entre la rivière Saigon et la rivière Dong Nai et les deux derniers sites sont localisés dans les canaux du bassin Saigon et Dong Nai.

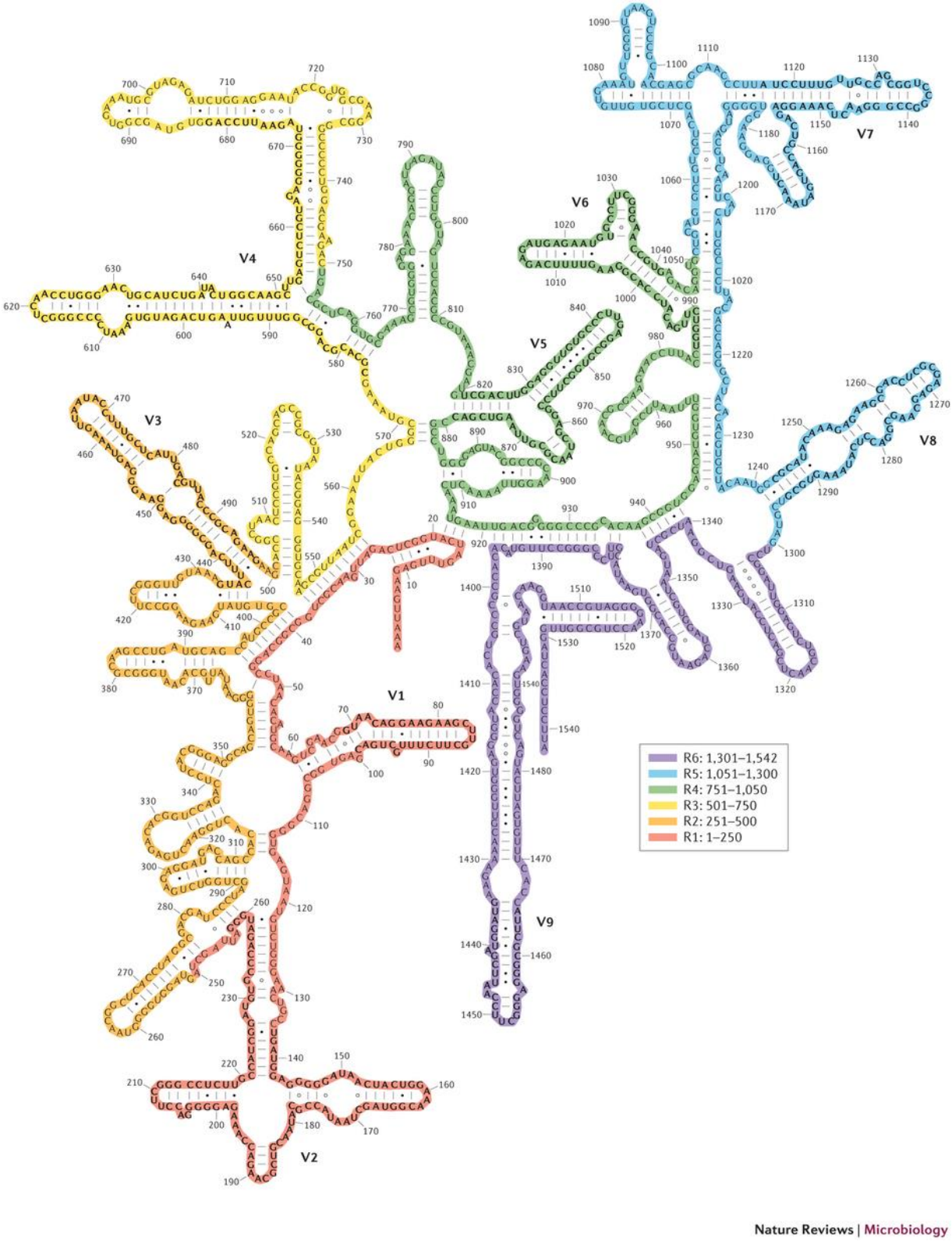


Figure B. Structure secondaire de l'ARNr 16S avec neuf régions hypervariables V1-V9 (en caractères gras) (**144**).

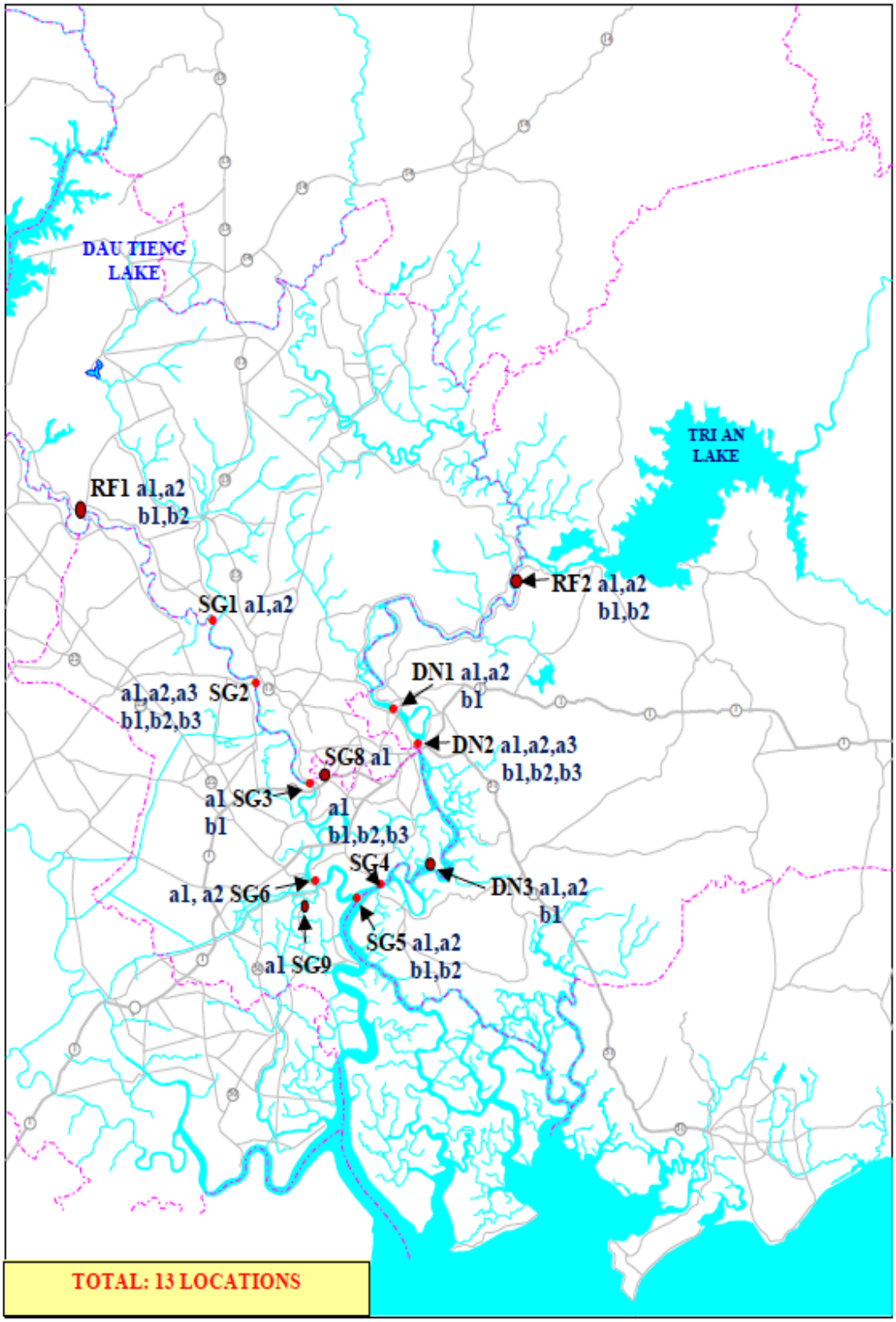


Figure C. Emplacement des 13 sites pour lesquels 42 prélèvements de sédiments ont été réalisés. Note : Pour les sites A et B, les échantillons ont été prélevés en triplicatas.

Dans un premier temps, nous avons déterminé les composés chimiques polluants la rivière. Des analyses effectuées au préalable ont montré que les hydrocarbures aromatiques polycycliques (PAH) sont les polluants les plus présents dans les rivières Saigon et Dong Nai. Les analyses des échantillons ont révélé la présence de polychlorobiphényle (PCB) ainsi que de différents métaux lourds. Afin de déterminer les différents PAH, nous avons analysé nos échantillons par chromatographie en phase gazeuse couplée à un spectromètre de masse (CG-MS). Les résultats ont montré que parmi 17 PAH standards, l'acénaphtène, le dibenz (a, h) anthracène, le benzo (j) fluoranthène et le benzo (e) pyrène n'ont pas été détectés avec un seuil de détection de 1 $ng.g^{-1}$ de poids sec.

Des analyses PCA avec différents types de traçage, appelées courbes GG (357), ont été réalisées sur les données de PAH des différents échantillons afin d'identifier des similitudes entre les différents sites des rivières Saigon et Dong Nai. Le résultat PCA GG montre que le groupement des PAH tels que l'anthracène, le fluoranthène, le pyrène, le benzo [a] pyrène, le benzo [g, h, I] pérylène, l'indéno [1,2,3-cd] pyrène, le benzo [a] anthracène + chrysène, benzo [b & k] fluoranthène et les PAH totaux sont corrélés avec l'échantillon SG8a1 provenant du canal (figure D). Ceci est en accord avec le fait que l'échantillon SG8a1 possède la concentration la plus élevée de ces composés de PAH et de PAH totaux (Tableau A). Le second échantillon prélevé dans le canal (SG9a1) présente également une forte concentration en anthracène, en fluoranthène, en pyrène, en benzo [a] pyrène, en benzo [a] anthracène + chrysène et en benzo [b & k] fluoranthène.. Les échantillons SG2a1 et SG2b1 provenant de la branche de la rivière Saigon montrent la plus forte concentration en pérylène comparé aux autres échantillons. La concentration en pérylène dans ces échantillons s'élève entre (807 et 630 ng.g-1, respectivement). Ces résultats montrent une pollution plus importante dans les canaux que dans les rivières Saigon et Dong Nai.

Les résultats obtenus lors de la PCA sont en accord avec des recherches menées sur la rivière Soltan Abad (Iran). Tout comme pour les rivières SG-DN qui passent par la ville de Ho Chi Minh, la rivière Soltan Abad passe par la ville industrielle Shiraz qui possède de nombreuses usines et industries (matériaux industriels et produits chimiques, caoutchouc et plastiques, objets en métal, etc.) provoquant la pollution de la rivière par des composés PAH. Les résultats obtenus lors de l'analyse des échantillons

des rivières SG-DN ont montrés que les concentrations en PAH dans les sédiments sont similaires à celles de la rivière Soltan Abad.

Dans une deuxième partie, nous avons comparé la diversité microbienne des 13 sites pour voir l'évolution des populations microbiennes en fonction de différents niveaux de pollution. L'étude a révélé que la population microbienne changeait de l'amont vers l'aval au niveau de l'OTU, du phylum et du genre après avoir traversé la zone urbaine très dense et la zone industrielle. De plus, comme nous avons démontré lors de notre première étude, les canaux du bassin versant des rivières SG-DN sont fortement pollués par de fortes concentrations de composés organiques (PAH) et possèdent différentes communautés bactériennes par rapport aux échantillons des rivières.

Dans notre première étude microbienne, nous nous sommes intéressé au nombre d'OTU dans nos échantillons. L'étude des populations microbiennes a montré que le nombre d'OTU dans la rivière DN était plus élevé que celui obtenu pour la rivière DN. De plus, le nombre d'OTU obtenu en aval de la rivière était plus faible que celui obtenus pour les branches des rivières SG et DN. Finalement, le nombre d'OTU obtenu dans les canaux était le plus faible de tous. Ces résultats sont en accords avec ceux effectués sur l'étude de la composition de la communauté bactérienne d'une rivière urbaine d'Afrique du Sud, dans la province du Nord-Ouest, qui a montré un nombre d'OTU plus faible en aval de la rivière (**319**). Une autre étude a montré que la composition de la communauté bactérienne des échantillons recueillis en amont de la station de traitement des eaux usées (STEP) était significativement différente de celle des échantillons en aval et des effluents de STEP (**320**). Finalement, une autre étude a comparé la diversité bactérienne des sédiments pollués et non pollués par des ignifugeants bromés et a révélé que la structure de la communauté bactérienne des sédiments pollués était différente de celle des sédiments non pollués (**287**). Ces différentes études sont en accords avec es résultats obtenus dans ce travail de thèse.

Dans notre seconde étude microbienne, nous nous sommes intéressé au phyla dans nos échantillons. Les communautés bactériennes des sédiments des rivières SG-DN sont dominées par les phyla *Proteobacteria, Chloroflexi, Nitrospirae* et *Acidobacteria*. Les Phyla *Bacteroidetes, Actinobacteria, Aminicenantes, Chlorobi, Planctomycetes, Verrucomicrobia, Spirochaetae* et *Firmicutes* sont apparus avec moins d'abondance parmi les échantillons de la rivière. Ces résultats sont en accords avec une étude, réalisée par Mark Ibekwe et al. (2016), portant sur la composition bactérienne

dans une rivière urbaine impactée par différentes sources de polluants et a montré que les phyla *Proteobacteria*, *Bacteroidetes*, *Acidobacteria* et *Actinobacteria* étaient dominants dans tous les échantillons de sédiments (**278**). Une autre étude dans les sols des parcs urbains de 16 villes chinoises représentatives utilisant le pyroséquençage a révélé que les 6 phylums dominants présents dans tous les échantillons étaient des protéobactéries, des *Actinobactéries*, des *Acidobactéries*, des *Planctomycètes*, des *Chloroflexi* et des *Bacteroidetes* (**279**). Des membres des β-Protéobactéries, des ε-*Protéobactéries*, des *Acidobactéries*, des *Bactéroïdes* et des *Verrucomicrobiens* ont également été trouvés dans la composition bactérienne d'une rivière urbaine dans la province du Nord-Ouest, en Afrique du Sud (**280**). La composition bactérienne des sédiments des rivières SG-DN partagent des caractéristiques similaires à celles d'autres sédiments urbains de différentes régions du globe (**278, 279, 280, 281**).

Dans une dernière étude microbienne, nous nous sommes intéressé au nombre genre de la population bactérienne de nos échantillons Les genres non cultivables de plusieurs familles ont été détectés dans des proportions élevées parmi les échantillons comme le genre *Anaerolineaceae*, *Nitrosomonadaceae* et *Nitrospiraceae*. Une étude de la population bactérienne dans les sols pollués par des métaux lourds, réalisée à l'aide de l'analyse de pyroséquençage de l'ADNr 16S, par Marcin et al. (2014), a montré que les quatre genres les plus abondants étaient des membres non cultivables d'*Acidobacteriaceae*, de *Gemmatimonadaceae*, de *Nitrosomonadaceae* et de *Xanthobacteraceae* (**293**). D'autres genres tels que *Nitrospira* étaient abondants dans les échantillons de sédiments des rivières SG-DN. Les membres de *Nitrospira* jouent un rôle important dans le processus de nitrification du cycle biogéochimique de l'azote. Lors de la première étape du processus métabolique, l'oxydant ammoniacal *Nitrosomonas* oxyde l'ammoniac en nitrite dans des conditions aérobies et le nitrite est oxydé en nitrate par *Nitrospira* (**294, 295, 296**). Des populations élevées de *Nitrospira* sont probablement causées par des concentrations élevées de nitrite dans les échantillons de sédiments. Les nitrites sont souvent utilisés comme inhibiteurs de corrosion dans les tours d'eau industrielle et de refroidissement (**297, 298, 299**). *Nitrospira* a été trouvé en tant que genre dominant en aval des stations d'épuration des eaux usées (STEP) dans la Seine (**300**). En outre, des bactéries nitrifiantes oxydant les nitrites (NOB) sont présentes dans les sédiments pollués de la rivière Niida (Hachinohe, Japon) ainsi que des bactéries oxydant l'ammoniac (AOB) (**301**). Des études de la Seine ont montré qu'il y avait accumulation de nitrite dans les stations aval (Poissy et Posses)

(**302**). Les concentrations d'azote total (ammoniac, nitrite et nitrate) de la rivière Saigon étaient plus faibles à l'emplacement amont de Thu Dau Mot (1,5 à 1,8 mg.l^{-1}), et plus élevées à l'aval du port de Nha Rong (2,4 à 3,2 mg^{-1}) (**303**).

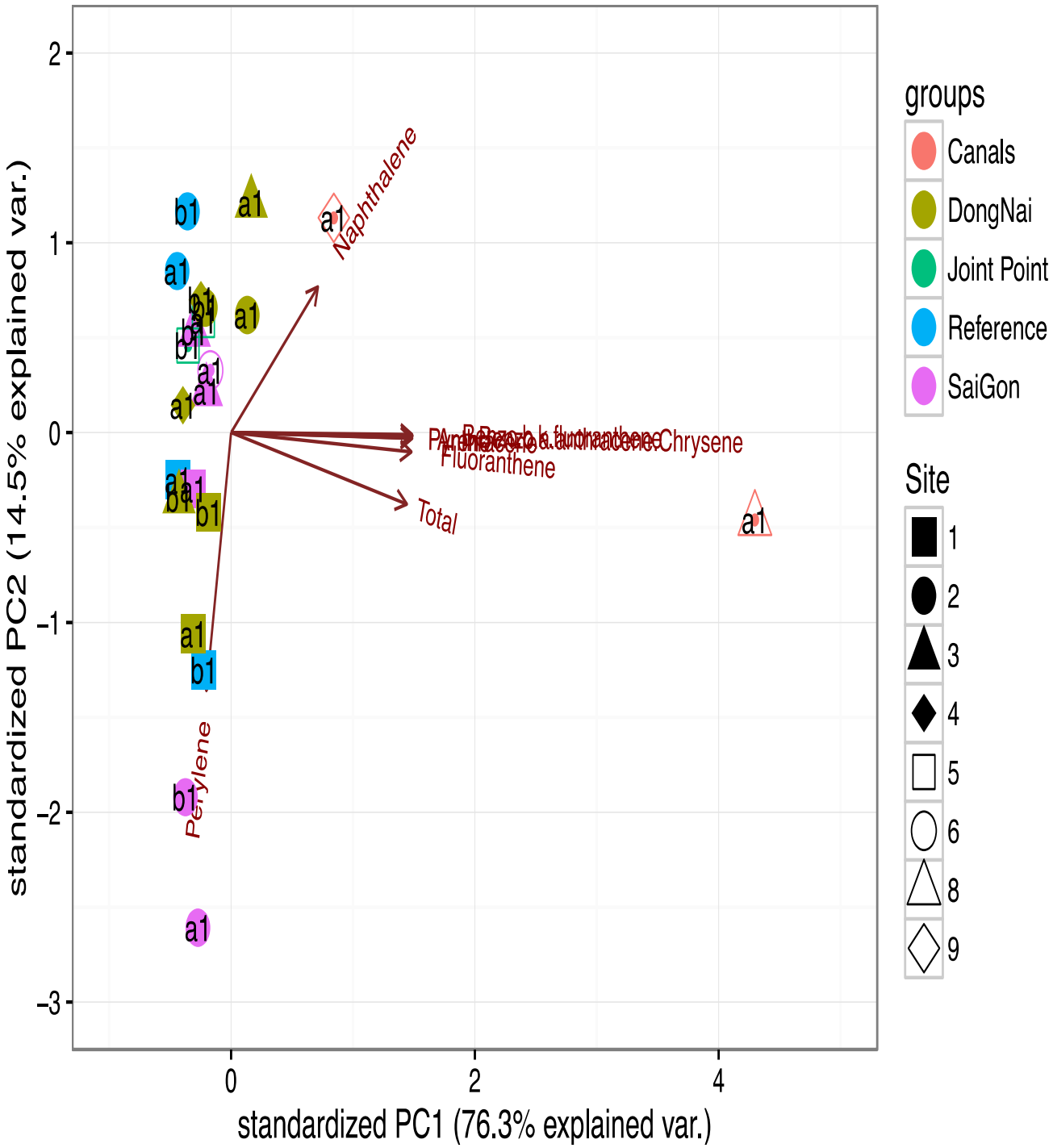


Figure D. Diagramme PCA GG des analyses chimiques (PAH) de 22 échantillons de sédiments (le premier échantillon du côté gauche (a1) et le premier échantillon du côté droit (b1) de 13 sites) du système fluvial SG-DN sur le deux premiers composants principaux (PC1 & PC2).

Les résultats obtenus peuvent montrer l'étroite relation entre la présence du genre *Nitrospira* et la présence de nitrates dans la rivière, indiquant les niveaux de pollution causés par les activités industrielles.

Lors de ce travail de thèse, nous avons analysé pour la première fois la pollution des rivières Saigon et Dong Nai ainsi que l'étude de la population bactérienne le long des rivières afin de déterminer l'impact de zones urbaines à forte densité et de zones industrielles. Les résultats ont montré que la pollution est plus importante dans les canaux ce qui a pour cause de limiter le développement de populations microbiennes. De plus, la population bactérienne est différente de celle en aval des rivières ce qui indique l'impact industriel et urbain.

Tableau A. Analyse des PAH (ng.g^{-1} poids sec) des échantillons de sédiments pour le premier échantillon du côté gauche (a1) et le premier échantillon du côté droit (b1) des 13 sites. Au total, 22 échantillons de sédiments ont été analysés et 17 composés de PAH ont été analysés (ND : non détecté avec un seuil de détection de 1 ng.g^{-1} en poids sec).

Location		Total PAHs	Naphtha-lene	Acenaph-thylene	Fluorene	Phenan-threne	Anthra-cene	Fluoran-thene	Pyrene	Benzo[a] pyrene	Benzo[g,h,I] perylene	Indeno [1,2,3-cd] pyrene	Benzo[a] anthracene +Chrysene	Benzo[b&k] fluoranthene	Perylene
SaiGon branch	RF1a1	415	87	ND	ND	ND	50	6	5	ND	ND	ND	ND	ND	267
	RF1b1	885	130	27	ND	ND	94	20	13	ND	ND	ND	ND	ND	601
	SG1a1	512	95	ND	ND	ND	49	27	26	ND	ND	ND	19	7	289
	SG2a1	1034	78	ND	ND	ND	78	34	22	ND	ND	ND	15	ND	807
	SG2b1	799	71	ND	ND	ND	47	20	19	ND	ND	ND	12	ND	630
	SG3a1	412	70	ND	ND	ND	85	44	61	7	ND	ND	25	27	93
	SG3b1	379	114	ND	ND	ND	45	34	54	ND	ND	ND	ND	ND	132
	SG6a1	458	99	ND	ND	ND	101	26	39	ND	ND	ND	33	22	138
DongNai branch	RF2a1	194	104	ND	ND	ND	36	7	6	ND	ND	ND	ND	ND	41
	RF2b1	228	130	ND	ND	ND	60	7	8	ND	ND	ND	ND	ND	23
	DN1a1	663	84	ND	ND	ND	65	16	23	ND	ND	ND	13	13	449
	DN1b1	719	141	ND	ND	ND	95	17	27	ND	ND	ND	14	ND	425
	DN2a1	750	184	ND	ND	ND	139	49	62	ND	ND	ND	33	27	256
	DN2b1	486	156	ND	ND	ND	86	14	21	ND	ND	ND	8	ND	201
	DN3a1	730	225	ND	ND	19	76	40	65	11	ND	ND	51	31	212
	DN3b1	387	64	ND	ND	ND	50	13	20	ND	ND	ND	7	ND	233
	SG4a1	343	87	ND	ND	ND	36	8	20	ND	ND	ND	14	9	169
	SG4b1	441	131	42	6	ND	73	15	27	ND	ND	ND	18	ND	129
Junction	SG5a1	465	126	42	ND	ND	101	16	24	ND	ND	ND	15	ND	141
	SG5b1	325	97	ND	ND	ND	78	11	15	ND	ND	ND	11	ND	113
Canals	SG8a1	3854	177	29	54	ND	712	720	702	237	122	138	454	350	159
	SG9a1	1093	198	ND	ND	ND	239	126	159	41	ND	ND	127	88	115