The ever-growing fascination of microbial ecologists with bacterial diversity continues to manifest itself in new techniques developed everyday for revealing the numbers as close to reality as possible. The initial consideration for such innovations is based on a rough estimation of the expected number of the population members from the ecological niche to be investigated. Same condition applies for studying the bacterial diversity of a given environmental sample. The choice of analytical strategy such as PCR primers and separation method for amplified DNA depends on this expected number (Wintzingerode *et al.*, 1997).

Considering the depauperate terrestrial biodiversity in Antarctica (Convey, 2010), the most favoured techniques for studying biodiversity remain DNA-based culture-independent methods including amplification of 16S rRNA from total genomic DNA extracted directly from environmental samples (Aislabie *et al.*, 2008). However, the most serious concern with culture-independent techniques is the potential bias introduced during various stages of the process (Wintzingerode *et al.*, 1997).

For starters, let us consider the first step of the process which is DNA extraction from the soil sample. For the current work, DNA extraction was carried out in the Rothera research station, Antarctica, and the extracted DNA was examined for purity on the beginning of the experiments. As suggested by the OD readings obtained from the spectrophotometer (Table 4.1), there is an unknown source of contamination in the DNA samples. The absorbance reading at A_{280} reveals how much protein or ssDNA/RNA is present. The A_{260}/A_{280} ratio value for all three samples is less than the tolerated range (i.e. 1.8-2.0). It is particularly lowest in DNA sample from VV site (0.59) which could indicate a high level of contamination by either ssDNA/RNA or proteins. However, A_{260}/A_{230} ratio, which indicates the organic matter presence in the samples, is below the optimal range of $\geq$1.5 only in DNA sample from LE site (0.44), suggesting traces of contamination by organic compounds in the sample.

As can be seen from Table 3.1, the LE sample collection site is associated with some vegetation while the other two soil samples lack this property. It can be inferred that the presence of organic matter in extracted DNA from LE soil sample may originate from here. DNA sample from VV site has the lowest value of A_{260}/A_{280} ratio, possibly suggesting a high level of ssDNA caused by shearing of DNA during cell treatment or protein contamination, or simply a lower yield of total DNA from the soil sample due to less number of bacterial cells compared to the other sites. This might be explained by the extremity of its climate condition since it is a more southern location than RO and LE (Table 3.1).

The second step in the process is PCR amplification of 16S rRNA gene from total genomic DNA. There is a lot of discussion revolving around the various problems arising when these methods are applied to environmental communities. Wintzingerode *et al.* (1997) listed some of the most common sources of bias in PCR amplification, some of which include inhibition effect by co-extracted contaminants, differential amplification, and formation of artefactual PCR products such as chimeric DNA. Although precautions can be taken to avoid these problems, most of the time they are inevitable.

As for the DNA samples from our three study sites, all samples showed A_{260}/A_{280} ratio values lower than optimal range of 1.8-2.0 and were suspected to have some co-extracted contaminants, mainly humic acids present in the soil samples. The humic compounds are the by-products of bacterial cells and are basically present wherever there is microbial growth. They can interfere with the activity of *Taq* polymerase and reduce the efficiency of PCR amplification. Differences in the activity of DNA polymerases in presence of an expected high content of humic substances cannot be underestimated (Wintzingerode *et al.*, 1997). A possible resolution for reducing the inhibition effects of contaminants in environmental samples is diluting the

original extracted DNA sample before adding in as template, which revealed satisfactory results (refer to Fig. 4.1).

The other reason for a low A_{260}/A_{280} ratio is protein contamination caused by either the proteins from bacterial cells themselves or originating from human or animals. In either case, the commercial kits available for DNA purification post-extraction should aim in reducing such contaminating factors to result in best quality DNA for the rest of the experiment. Keeping in mind that small sample volumes are prone to nucleic acid loss during purification and have to be carried out with caution. In brief, low DNA yield from VV might be resulted from presence of protein contamination, or it can also be interpreted as that the VV contained naturally less DNA than the other two locations due to more extreme climate condition.

The template for PCR amplification is a heterogeneous mixture of DNA molecules, but the amplification efficiency is not always equal for all templates. As a matter of fact, some of the most abundant templates in the mixture tend to self-anneal in the late stages of amplification causing an inhibition of polymerase reaction (Acinas *et al.*, 2005). The outcome will be a distorted view of the genetic diversity from the source environment since the more abundant types will be represented in lower quantity in the final PCR product. On the other hand, less abundant genotypes in the community DNA have a tendency to be overrepresented. From here on, a non-realistic proportion of the 16S rRNA genes from different bacteria may be presented. By cloning the product into cloning vectors to separate out the resulting genes and further sequencing them, the results may show a higher quantity of those genes which are actually less common in the natural habitat. At the moment, this is a major source of bias in PCR-based techniques used for community structure studies and should be taken into consideration when interpreting the results (Wintzingerode *et al.*, 1997).

Despite this shortcoming, PCR amplification of 16S rRNA gene remains the top method of choice when it comes to studying environmental bacterial diversity due to its ease of use and the available established database as reference. In comprehension, since the studies being compared all use the same technique, the results are comparable assuming they all share the same level of bias at identical stages of the process. Therefore the bias can be overlooked for the sake of simplicity of interpreting the results and data analysis.

When applying restriction fragment length polymorphism (RFLP) analysis as distinguishing method, the main concern as possible source of error in the final number of individuals in phylotypes grouping might result from the RFLP fingerprint patterns. The choice of restriction enzyme (RE) employed is the key factor in the final results and theoretically, the more RE used in RFLP analysis, the more accurate the results will be. Nonetheless, the use of two RE have been shown adequate for analysis of species composition in complex communities (Wu *et al.*, 2006). In this case, the combination of two enzymes generates a more diverse banding pattern as opposed to just one RE, and thus a finer resolution for species level differentiation. However, at some point the banding patterns can become too complex to be detected by a routine agarose gel electrophoresis. At these instances, two fingerprints with absolutely identical banding patterns on an agarose gel might be in fact totally different under closer inspection by a more accurate technique such as fragment analysis. Not to mention the possibility that different species can reproduce similar fingerprint patterns given certain RE.

In general, a drawback of community analysis based on these techniques is that a population must represent at least 1% of total community to be detectable in a fingerprint. This means that less abundant taxa are frequently left out or grossly underestimated in such studies (Smalla, 2004). At any rate, differentiating between bacterial types using RFLP saves time and cost, but has the risk of false categorization

due to similarity in banding patterns. In our study, we used 1.5% agarose gel to separate the resultant bands. Obviously, the resolution achieved this way is not ideal for accurately distinguishing between very close banding patterns. As an example, phylotype A, which was the most commonly occurring phylotype, had a banding pattern with two bands at 200 bp and below. There were several cases where the fingerprints on the stained gel resembled this double-band pattern, but in some the second band would appear thicker suggesting a cluster of multiple bands. This could have been confirmed by another gel system, e.g. polyacrylamide gel, which is more suitable for separating smaller size DNA. As an alternative, several individuals within a phylotype group can be sequenced for confirmation, but was not feasible here due to budget limitation.

After conducting the final list of all phylotype groups and their relative abundance for each clone library, it is evident the diversity is lowest in soil sample from Rothera (25 phylotypes) and highest in Léonie Island (35 phylotypes). Viking Valley comes in the middle with 31 phylotypes. Even though the numbers seem very close to each other and the percentage of phylotypes shared between three sites is around 60%, the three sites are not identical as shown by a non-metric MDS (refer to chapter 4, section 4.8.3.3). Furthermore, the sampling effort for LE site was not exhaustive according to the corresponding collector's curve (Fig. 4.8.a), suggesting the presence of a few more phylotypes to be discovered in this site.

As mentioned earlier, the difference in the pattern of biodiversity along the three study sites does not appear to be random and can be explained according to environmental conditions in each sampling site. In general, increasing latitude has been linked with a decline in bacterial diversity (Yergeau *et al.*, 2007). Léonie Island and Rothera sampling sites are located at equal latitudes (67^{o}S) and are expected to show similar climate condition with a similar pattern of biodiversity. However as mentioned earlier, the presence of vegetation in the LE site serves as an advancing factor in

increasing the microbial profile of the soil. As it is known, vegetation of the land helps decrease the harshness of sever Antarctic climate by providing enhanced moisture and thermal maintenance (Yergeau *et al.*, 2007). This might be one of the reasons as why two geaographically identical locations harbour different sets of biodiversity. But it still does not explain why at a lower latitude, RO supports less number of bacterial diversity when compared to Viking Valley with higher latitude and hence, more severe climate conditions.

Rothera Point (RO) was sampled within the vicinity of Rothera Research Station where high anthropogenic influence was expected. For instance, fossil fuels used for transportation and power generation have been shown to cause petroleum hydrocarbon contamination of soils, which may result in enrichment of culturable heterotrophic bacteria and hydrocarbon-degrading microorganisms (Saul *et al.*, 2005). This means that manipulations of the natural environment caused by human activities might result in creation of a selective environment in which only certain microorganism types are able to adapt and survive. This may be another reason for a lower biodiversity in RO compared to the other two sites.

Léonie Island and Viking Valley clone libraries revealed higher diversity with 35 and 31 phylotypes, respectively. It is not surprising to find the LE site sample to harbour higher diversity of all three studied sites, given the fact that the ice-free land supports a diverse and dynamic vegetation profile. Viking Valley, on the other hand, is nearly barren and constantly covered with permafrost, and is located at higher latitude (71^{o}S). Nonetheless, it was found the second most populated site out of the three, with some unique phylotypes that were not observed in any of the other two sites as well as the *Firmicutes* phylum which was exclusive to this site alone.

Once again it must be reminded that the quality of the results obtained from comparative 16S rRNA gene sequence analyses is dependent on the dataset used. With all the full and partial sequences of cultivated microorganisms and environmental clones released in public databases, the numbers do not come anywhere near the expected microbial diversity. Consequently, many environmental samples exhibit low sequence similarity to known sequences due to either novel origins or unavailable cultured representatives and in some cases, poor sequence quality (Wintzingerode *et al.*, 1997).

Another important consideration in assessing bacterial diversity is whether the 16S rRNA sequence recovered from PCR-mediated approach belongs to a living organism or it has originated from a dormant or dead cell preserved in the soils. Nucleic acid molecules, particularly DNA, can prevail in the environment for long periods of time, especially in cold temperatures of the Antarctic soil. Therefore, it is possible that a percentage of observed diversity could have derived from preserved DNA rather than from viable bacterial cells. Wintzingerode *et al.* (1997) suggest the use of 16S rRNA in situ hybridization as opposed to PCR-mediated approaches since they are able to detect actively growing microorganisms with a sufficient content of target molecules.

The major phyla identified in the current work are characteristic soil bacteria including *Bacteroidetes*, *Proteobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Cyanobacteria*, *Actinobacteria*, and *Firmicutes*. Although the proportions of relative abundance in each phylum differ from site to site, all phyla were well represented in the three sites with a few exceptions. For instance, *Cyanobacteria* was not observed in VV site sample, which might be explained by the low water availability in soil necessary for growth of this type of bacteria (Smith *et al.*, 2006). On the other hand, *Firmicutes* which are capable of forming endospores and show high resistance to desiccation stress (Dorland, 2008), were only present in the clone libraries from VV soil sample.

As it is evident from the results obtained in this study, bacterial diversity throughout the continent is highly influenced by environmental factors and therefore, species richness is more restricted than temperate regions. This, however, is usually compensated by a wide range of taxa making up a diverse gene pool with high metabolic flexibility to manage most essential functions of Antarctic ecosystems (Wynn-Williams, 1996). What is not completely understood by 16S rRNA gene analysis carried out in molecular approaches such as the current study is a full view of functional importance of soil bacteria. With all the bits and pieces of data inflow towards the pool of information on bacterial diversity, there is still little known about the relationship between phylogenetic diversity and functional diversity. Therefore, new approaches must be sought after to link phylogeny with function by merging information about both cultured and uncultured microorganisms, along with species abundance, nutrient cycling, signal molecules, antibiotics, and other small diffusible compounds (Handelsman, 2004).

5.1 Conclusion

Soil bacterial diversity from three sites on Antarctic Peninsula showed low biodiversity compared to temperate soils as expected. 16S rRNA gene clone libraries representing soil samples across Rothera Point (RO), Viking Valley (VV) and Léonie Island (LE) sites revealed 25, 31 and 35 phylotypes in each clone library, respectively. Out of total 46 phylotypes found, 15 were shared by all three sites, while 8 phylotypes were exclusive to LE, 6 were occurring only in VV, and 2 were found only in RO. Pair-wise comparison between the sites revealed 7 phylotypes being shared by VV and LE, 5 by RO and LE, and 3 by RO and VV. A trend of environmental influence on bacterial diversity might be evident since the three sites have characteristically different environments, with VV site having the most extreme climate compared to LE with some vegetation and a somewhat less stressed climate. Between the three sites, RO showed the least number of phylotypes, possibly suggesting a negative impact of environmental manipulations on bacterial diversity caused by human activities. LE site revealed the highest diversity with few more phylotypes left undiscovered. All discovered phylotype sequences showed close relationship (i.e. ≥95% similarity) to sequences in GenBank with origins of Arctic's and Antarctic's.