Soil microbial diversity across different agroecological zones in New South Wales

Vanessa Pino V.

Faculty of Agriculture and Environment

The University of Sydney
New South Wales, Australia

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Summary

A synergistic relationship between *soil diversity* (pedodiversity) and *soil microbial diversity* (biodiversity) seems axiomatic. Soil microbial communities regulate the essentials biogeochemical cycles (e.g. N and C cycles) on which rely the functioning and services of soil ecosystems (e.g. food production, C storage). On its own, the soil matrix provides and regulates living conditions by which these microbial communities are structured and functional. Nonetheless, the multi-dimensionality of this *pedodiversity-biodiversity* relationship is still poorly understood. A better insight into this synergy would enable us to quantify/qualify and so sustain, protect, and hopefully improve, those processes underpinning soil functioning (e.g. promote N-fixation and C sequestration). From a soil scientific perspective, we hypothesise that the structural and functional extent of soil microbial communities relies on multivariate soil units (e.g. soil horizons, profiles, classes) instead of any single discrete 'environmental factor' (e.g. soil pH, precipitation) and; *vice-versa,* the structure of these communities can become a well-defined biological property of these soil entities. We began exploring the multidimensional disposition of this biotic-abiotic functioning by modeling the biogeographical patterns of soil microbial communities - richness and diversity - using biomolecular sequencing, pedometrics, and digital mapping approaches.

As a first exploratory analysis, this particular thesis evaluated local diversity (α-diversity) of *bacteria*, *archaea* and *fungi* communities using a latitudinal ($NS_{\text{-transect}}$) and a longitudinal ($WE_{\text{-transect}}$)

transect of about 900 km each across New South Wales (NSW), Australia. Along these transects, soil ecosystems were sampled at 48 different locations from paired conserved (e.g. woodlands) and disturbed (*e.g.* cropping) ecosystems. Afterward, soil biophysicochemical attributes were estimated using 16SrDNA and ITS metabarcoding (11,557,499 sequences; 423,740 OTUs) and pedometric approaches (19 soil properties; 13 environmental covariates).

The microbial structural patterns, abundance and local diversity (α-*diversity*), were assessed in relation to the soil physicochemical properties using both linear and quadratic associations and other multivariate analysis (e.g. PCA, bootstrap regression modelling and, mapping). This enabled the spatial prediction and mapping of the three microbial kingdoms at a resolution of 1#km across all of NSW. Our maps showed soil microbial diversity, richness, and abundance following a combination of soil and environmental attributes in which western NSW has of higher diversity compared with eastern NSW. Despite this gradient, *fungi* and *archaea* were consistently lower and higher in *Vertosols (Australian Soil Classification System)*, respectively, whereas the distribution of *bacteria* is less clear. Our results suggest that the structure of microbial communities is intimately related with most physicochemical soil attributes but this association, whether linear or not, varies not only upon one single soil properties but a group of them. At the same time, the extent and direction of these relations vary accordingly the different microbial taxa (e.g. by phylum). Therefore, microbial diversities are more consistent with grouped features defining soil entities (*e.g.* horizons and profiles classes) rather than on individual soil attributes. Our further work will include more evidence for these conclusions by analysing microbial and pedological dissimilarities (*e.g. β-diversity*) in a multiscale approach.

Acknowledgements

"Mapu mew iñchiñ ta ngen-ngenkülelaiñ,
iñchiñ may ta mapu ngeiñ" (mapundungun)

La tierra no es nuestra.
Nosotros somos de la tierra
(proverbio mapuche).

There are two immediate thoughts that come to mind after completing this Ph.D. dissertation. First, that the magical world beneath our feet never ceases to impress me and it seems, after all this hard work, I must sincerely be in love with. My second appreciation is that, more than ever before, I know how much I do not know that still challenges my curiosity – which I think is a very valuable motivation for a scientist.

At the end of this journey, I would like to express my sincere gratitude to my supervisor Alex McBratney for his guidance but mostly for having given me, since the start, the freedom to decide to work on soil biology and then the confidence for letting me be a part of such a beautiful project. Many thanks to Neil Wilson and Rosalind Deaker for their co-supervision and initiation into soil microbiology and molecular techniques. Thanks Neil for your confidence on my laboratory skills. I also would like to thank all my friends, colleagues and technical stuff that gave a bit of their time to help me when carrying out fieldtrips, laboratory work, data analysis or simply those fun Fridays sharing some beers. Uta, Kanika, Yolima, Phil, Lori, Ivan, Stacey, Liana, Ed, Patito, Claudia, Damien & Poppy, Budi, Jay, Olivier, Jessica, Mirja, Kasia, and many more.

Many thanks to my Australian family and loved ones that brought laughter, love and fun to my days away from my Chile. Chalmers family (Flo, Paul, Sofi, Maryam, Jose, Clara), Sandrita & all the CCFO dancers (I leap into heaven when dancing with you), Patito, Romi & The wonderland home, and Claryto (que no alcanzaste a editar este pedacito de texto). A special thank you to Mario Fajardo, we absolutely did this together, thanks for your unconditional friendship, love, support and *cafesongos* every day.

Finally, I would like to dedicate this work to my family. To my dad, my mom and my Grandma, Irene, Jose Luis and Normita, thanks for always encouraging me to be independent and strong; to my siblings Muriel, Belen, Moises and Zaira, love you all. To my two big families; my many uncles, aunties, cousins and nieces; and to my Chilean friends who are always looking after me; childeltas, callejeras, las carmelas, georgies and others.

Muchas gracias a todos!

Table of Contents

Chapter 1

Soil Biodiversity & Pedodiversity: Synergies and Parallels

Outlines

Understanding the spatial soil biodiversity-pedodiversity relationship can provide useful information when understanding soil microbial structural patterns in different environmental gradients. In this research, this relation has been demonstrated by providing a robust and deep analysis of the soil microbial structural diversity in relation to the soil physicochemical characteristics of a longitudinal and latitudinal environmental gradient across NSW, Australia. This chapter provides the foundation for this study, informing the relevance of soil microbial studies in the new era of modern microbiology and pedometrics science.

Summary

This chapter provides a general perspective on how this study contributes to characterizing soil microbial communities (biodiversity) in the context of their soil environment (pedodiversity). Apart from each other, soil microbial ecology and soil science have been recently realised important advances in the study of soil biota and soil physicochemical properties but still these disciplines did not converge their knowledge to a common and complete understanding. Heretofore, soil scientists have developed modern approaches predicting the distribution of soils around the globe as well as its physicochemical arrangements but barely taking into consideration the soil biological components. On its own, ecologists have increased our concern about soil microbiota putting a considerable effort these days on understanding their structural/functional diversity and biogeographical patterns. We advise that soil heterogeneity (pedodiversity) is already being predicted on a global scale on the basis of soil formation factors/processes for which these pedological approaches might indeed define the spatial patterns of soil microbial communities. This notion lies behind the framework for our aims and the methodological advantages in this investigation.

In this chapter, we also examine the benefits of this new understanding. Indeed, the prediction of soil microbial patterns in the entire context of their soil environment would able not only to qualify but quantify their influence upon biogeochemical processes in the soil functioning (e.g. for ensuring food production). Any advance towards the quantification of soil processes and functioning enables its assessment for decision-making and regulatory policies for soil protection (Soil Security).

SYNERGIES AND PARALLELS OF SOIL BIODIVERSITY-PEDODIVERSITY

Soil biodiversity has been widely documented in relation to a set of soil physicochemical properties but never to the fine extent of pedodiversity or 'soil heterogeneity' – i.e. the diversity of soil properties, materials, and classes (McBratney et al., 2015). Over time, synergies and parallels of this pedodiversity-biodiversity relationship have been 'partially' demonstrated from a different perspective, e.g. pedology and bioecology.

From a pedological view, soil organisms - or 'soil fauna' or 'soil biota'– have been widely identified for nearly century as 'one of the five fundamental soil-forming factors' contributing in the soil genesis - together with *climate*, *topography*, *parent material* and *time (*Jenny, 1941). Furthermore, the soil biological component is an essential contributor to the soil environment by influencing soil physicochemical properties (e.g. soil fertility and soil structural features) and so soil functional performance (e.g. soil capability for growing plants affecting both plant diversity and food production).

For example, regarding soil chemical attributes, soil organisms are intimately related with nutrient cycles (e.g. C, N, P, etc.) and, therefore, they contribute by conditioning *soil fertility* status. For example, decomposer organisms (e.g. bacteria, fungi, and earthworms) work on breaking down organic materials (e.g. plant, animal and microbial residues) with the subsequent delivery of new-fangled nutrients and organic substances (e.g. organic C, humus) in the soil

medium; or others such as nitrogen fixers (e.g. diazotrophic bacteria) and nitrifying organisms (e.g. ammonia-oxidizing archaea and bacteria) work on the N cycle, in which, the former fix atmospheric N in the soil (the only natural form of nitrogen fixation identified so far) and, the latter transforms this nitrogen in other nutritional forms such as ammonia, nitrates, and nitrites. Accordingly, the dynamic of these nutritional conditions influences others soil properties – e.g. cation exchange capacity, pH, electric conductivity, etc.

The influence of soil biological components over *soil physical attributes* has also been widely documented. Certainly, one of the most recognized contributions of soil organisms is their role during the formation of soil aggregates and, therefore, their influence on the *soil structural* properties as well as on other closely related ones such as *soil aeration (porosity)*, *hydraulic conductivity,* etc. For example, from a very simplistic view, this aggregation process involves fungal hyphae (particularly from arbuscular mycorrhizal fungi) and bacterial exudates (e.g. polysaccharides) acting as agents for forming soil macroaggregates (soil agglomerates larger than 250 μm) by binding and enmeshing soil microaggregates (soil agglomerates smaller than 250 μm) – this is a complex phenomenon that involves much more processes detailed in the popular model of 'aggregate hierarchy' (Six et al., 2004). Macroorganisms - particularly earthworms -, also contribute in this aggregation process by making it faster and increasing the rate of aggregation – e.g. mineral and organic soil particles are mixed up and forming aggregates when passing through the gut of a worm -. At the same time, their burrows improve porosity and so the diffusion and movement of water and gas (e.g. carbon dioxide and oxygen) maintaining them in balance. Both structural stability and gasses (water, oxygen, and CO_2) balances are critical

factors for plants rhizosphere environ (e.g. nutrients absorbance) and therefore for 'food production'.

More recently, the soil biological component, i.e. 'soil biodiversity', has been recognised as valuable since it plays a primary role in soil functioning. This is becoming a major concern these days as more evidence accumulated suggesting that soil biodiversity loss and simplification of soil community composition impair multiple ecosystem functions. For example, Wagg et al., in 2014 confirmed the negative effect soil biodiversity loss over plant diversity, decomposition, nutrient retention, and nutrient cycling. Later, Delgado-Baquerizo et al., in 2016 provided empirical evidence showing that any loss in microbial diversity will have a negative impact on climate regulation, soil fertility and food and fibre production in terrestrial ecosystems.

From one to another direction so far the direct influence of soil biota over soil physicochemical attributes has been revealed. But *vice-versa*, soil physicochemical attributes has been also broadly exposed as primary factors controlling soil organisms and, particularly, the soil microbial communities structure (i.e. composition, abundance, diversity, etc.) highlighting the synergetic character of this pedodiversity-biodiversity relationship.

For example, soil attributes (e.g. soil pH) have been found to dictate the structural patterns of soil microbial communities across space, at both regional and/or local scales (Cao et al., 2012; Dimitriu and Grayston, 2010; Izquierdo and Nüsslein, 2006; Lauber et al., 2008; Xiong et al., 2012); and time, when soil composition has meant a critical factor controlling the stability of

the rhizosphere microbiota in microbial successional studies (Tkacz et al., 2015). Nevertheless, from our pedological point of view, soil microbial communities for being immersed in the soil matrix are not merely in relation to a single soil attribute or property and, definitely, these tiny communities are shaped by a group of them, instead. Such group of attributes represents a certain degree of 'soil heterogeneity' (pedodiversity) that usually delimitate, differentiate and classify specific soil entities (e.g. horizons, classes, orders). A co-spatial relationship between soil microbial structural aspects (diversity, abundances, composition, dissimilarities) and pedodiversity has not yet been analyzed at regional scales - either horizontally (e.g. classes) or vertically (e.g. horizons, profiles). There are mostly local studies (Baldrian et al., 2012; Huang et al., 2014; Rime et al., 2015; Steven et al., 2013; Vos et al., 2013) and many others at large scale but focused on other environmental gradients (e.g. altitude) (Nunan et al., 2002; Yasir et al., 2015; Zhang et al., 2013). Therefore, soil microbial community patterns have not strictly been described in the fine context of soil heterogeneity, i.e. as a pedodiversity-biodiversity compound. The fact that microorganisms do not merely respond to a single soil property/attribute but instead to a 'soil heterogeneity gradient', was analyzed from a micro-scale perspective by Vos et al., (2013). These authors stated that '*the fine heterogeneity of soil results in a complex mosaic of gradients selecting for or against bacterial growth*'. In pedology, the turnover of this '*fine heterogeneity*' discriminates among soil diagnostic horizons, types, classes, etc. (Fajardo et al., 2016a; Hartemink and Minasny, 2014); yet we are missing a proper characterization of the soil microbial component in spatial correlation with these soil variations.

We argue that even though soil physicochemical properties and soil type have been found determinant factors controlling soil microbial community's structures at local (Garbeva et al., 2004; Girvan et al., 2003) and larger scales (Ranjard et al., 2009) there is still a lack of evaluation within a pedological context, e.g. soil microbial diversity distribution per soil taxa or per horizons, i.e. in the genuine and full context of a soil gradient. In addition, we also argue that when soil microbial studies emphasize their analysis on individual soil properties (e.g. clay, organic carbon, etc.) (de Gannes et al., 2015; Rousk et al., 2010) there is a valuable and important amount of information that is not being included that obscure our understanding of the multidimensional relation between soil biodiversity 16and pedodiversity. The most clear evidence to support our hypothesis is that many of the large-scales soil microbial studies (Chong et al., 2012; Griffiths et al., 2011; Lauber et al., 2009; Nemergut et al., 2011; Ranjard et al., 2013) demonstrate similar geographical patterns as those recognised in the geographic distribution of soils around the globe for more than a century in pedological studies. This suggests that a co-spatial relationship between aspects of soil biodiversity and pedodiversity is to be expected (Ibáñez and Feoli, 2013; Vos et al., 2013). Reasonably, the great majority of the investigation about soil microbial geographical patterns at wider scales aim to respond more ecological questions/hypothesis such as (i) is the biogeography of microorganism and macroorganisms similar, or (ii) does soil microorganisms diversity increase or decrease toward tropical/poles areas (Martiny et al., 2006); and many other related to determine patterns and processes of microbial assembly (Nemergut et al., 2013). Indeed, there is still an ongoing discussion about the controversial hypothesis attributed to Baas-Becking (1934) who stated '*everything is in everywhere*, *but the environment selects*'. This statement, from our point of view it would be '*what*

if…. pedodiversity select…' By saying that, we assume that whatsoever is there '*soil microorganisms have spatial patterns by following pedological gradients*' – our hypothesis in this study.

Why we need to explore soil microbial geographical patterns

By giving responses or not to ecological questions, the understanding of spatial microbial patterns enables us to elucidate their responses to environmental gradients. By gaining this knowledge we may be able to manage, improve and protect the role of soil microorganisms working for soil ecosystems services.

The synergistic relationship between soil functioning and microbes has been widely recognized. From the one side, there is an active participation of microbial communities behind every service provided by soil ecosystems (e.g. food production) (Nesme et al., 2016). From the other side, soil gradients (pedodiversity) is a primary factor controlling soil microbial communities (biodiversity) not only their spatial distribution across space (vertical and horizontal) and time (e.g. controlling microbial stability in the microbial succession process) but also their structural and functional diversity.

The valuable performance of soil microbes working for important biogeochemical processes importantly defines soil functioning and services. This fact has opened warning questions that must be answered such as if the reduction of soil biodiversity affects soil functioning

performance (M.J. Swift M. van Noordwijk et al., 2004; Singh et al., 2014). Certainly, *soil biodiversity* has a crucial role contributing to soil ecosystems services. Soil ecosystems services has been defined into four categories, i.e. *supporting*, *provisioning*, *regulating* and *cultural services*; and within them, *soil biodiversity* has an active involvement for *supporting* (e.g. nutrient delivery, soil formation, ensuring gene pool and biodiversity conservation), *regulating* (e.g. regulating of major elemental cycles, disposal of wastes and dead organic matter) and, *provisioning* (e.g. food, freshwater, genetic resources) as detailed below.

Soil microbial role in soil ecosystems services

The importance of soil functionality beyond the concept of '*ecosystem services*' emerged in the early 80s (Millennium Ecosystem Assessment et al., 2005). Since then, all benefits provided by natural ecosystems to human welfare are framed under this concept. For this reason, '*ecosystems services*' are being carefully well defined and classified for their economic valuation (De Groot et al., 2002). Once it has been defined an economic valuation for a determined *ecosystem service,* we allow making decisions for their protection (e.g. policy regulations). Nevertheless, there is still and undefined framework for *soil services* even though these goods are widely recognised - i.e. *provisioning of food, freshwater, fibre, fuel and genetic resources; regulating air and water quality along with climate sustainability; cultural services as maintaining the heritage and protection of the complete ecosystem; and supporting biogeochemical processes for nutrient cycles even the process of soil formation* (De Groot et al., 2002; Fisher et al., 2009; Nahlik et al., 2012; Robinson et al., 2013). Indeed, McBratney et al., (2014) argued that soil functioning is a common denominator for all terrestrial ecosystems (e.g. agriculture, forests, grasslands, deserts, and urban areas). And

according to this fact, security of food, water, energy, climate stability and biodiversity services are depending on soils functioning introducing the importance behind Soil Security (McBratney et al., 2012).

Beyond any of this purposes, i.e. either to unify a framework for the economic appreciation of *soil ecosystem services* or keep evolving the Soil Security concept, we need to determine soil functioning on which pedodiversity-biodiversity relationship is a key factor. Towards to reach this aim, the monitoring, measuring, mapping and modeling at different spatiotemporal scales are useful to come to a joint a consensus (Fisher et al., 2009; McBratney et al., 2015, 2014).

Since microbial communities have quick responses to environmental changes, they have resulted in an efficient tool for soil monitoring programs. They can be used as a research tool when estimating environmental changes owing to climate, contamination or degradation/erosional processes (Buckland et al., 2005; Hazen et al., 2013). For example, the estimation of microbial 'resilience[i]' and 'redundancy[ii]' was empirically useful for predicting rates of processes in ecosystems modeling facing global changes (Allison and Martiny, 2008). Similarly, Hu et al., (2011) demonstrated that the monitoring of soil microbes helped to estimates the effect of long-term fertilization in agricultural soils. As a result, this investigation highlighted that organic amendment plus balanced fertilization of N, P, and K, promoted soil microbial functional diversity and thus enhanced crop growth and production.

[i] *The rate at which microbial composition returns to its original composition after being disturbed.*
[ii] *the ability of one microbial taxon to carry out a process at the same rate as another under the same environmental conditions.*

One of the most documented roles of microorganisms has been their participation in essential biogeochemical processes. Food production and climate mitigation, for example, are services strongly dependent on carbon (C) sequestration and nutrient cycling. They involved two processes greatly governed by soil biota: formation of stable and labile carbon and processing of nutrient pools (Schulz et al., 2013).

The key participation of soil microbes for soil functioning performance might be related to the extent at which they are present in soil ecosystems, i.e. level of microbial biodiversity hosted by soils. Soil habitats probably contain the greatest microbial diversity of all the environments on Earth (Griffiths et al., 2004). This biodiversity even exceeding that of aquatic systems (Torsvik L., 2002). Thus, soil microorganisms, i.e. bacteria, archaea, viruses, protists, and fungi, constitute the most ubiquitous, diverse and abundant group of organisms on Earth (Fuhrman, 2009; Ranjard et al., 2003; Singh et al., 2014).

New applications of uncultured microorganisms and especially from the rare communities is becoming of special attentions for being a great source of genetic material and functional microbial diversity (Lynch and Neufeld, 2015; Reid et al., 2011). Bioprospecting[iii] ranges from human health (antibiotics), industrial, agricultural and environmental applications. Reid et al., (2011) in a way to explain the importance of rare biosphere stated that '*even though when have accessed less than 1% of the genetic diversity of life on Earth our antibiotics, our evolutionary perspective, our biotechnology, are all based on only a small portion of the potential diversity on the*

[iii] *The screening of biological systems (for example, genomes or ecosystems) for novel components of industrial, commercial or scientific value* (Lynch and Neufeld, 2015).

planet. We need to explore the other 99%.'

In summary, the roles of soil microorganisms encompass multiples areas but their contribution beyond soil functioning to ensure soil ecosystems services can be considered the most important ones since soils are vital but non-renewable resources at the human timescale (Banwart et al., 2015). By modeling the biogeographical patterns of soil microbial communities we gain understanding on what is making these tiny communities to change, improve or reduce their performance. For example, these days, we are able to guide field conditions for specific symbiotic microbial communities (e.g. legume/rhizobia symbiosis; Howieson and Dilworth, 2016) but this management is not possible for free- living microorganisms, especially rare ones. The challenge is that as such we provide guidelines for sustainable agricultural practices depending on agroecological conditions we could also regulated practices for soil microbial conditions.

The opportunity for pedodiversity-biodiversity large input analyses

The opportunity to explore the co-spatial relation of soil microbial and physicochemical attributes at larger scales resolution was not possible if not until recently. Clearly, the order of magnitude of these analyses required both major technological and economical efforts. Nonetheless, nowadays we possess the technological (instrumentation) and methodological (e.g. modeling abilities) advances in both soil (pedometrics) and biological (biotechnology) sciences which make this type of studies a plausible task.

Modern soil sciences encompasses historic advances for which this scientific discipline has derived in different branches (e.g. pedometrics, digital soil morphometrics). Pedometrics is a soil science branch which by applying quantitative methods aims to gain a better understanding of the soil as a phenomenon that varies over different spatiotemporal scales (Burrough et al., 1994). Similarly, digital soil morphometrics applies tools and techniques for measuring and quantifying soil profile attributes and deriving continues soil depth functions[iv] (Hartemink and Minasny, 2014). Modern soil measurement techniques provide access to high quality quantitative continuous information to picture the soil variation (e.g. Vis-NIR soil profiles) as also add more attributes (e.g. soil aggregation) for our analysis (Fajardo et al., 2016b).

On its own, microbiological sciences emerged to modern research capabilities as soon as new technologies started being adopted. Thus, microbial culture-independent methods (CIMs) with the inclusion of biomolecular laboratory techniques (e.g. PCR technique, metabarcoding), High-Throughput Sequencing Technologies (HTS) and bioinformatics tools enabled the possibility of high resolution microbial identification based on their DNA information (Logares et al., 2014; Reuter et al., 2015; Schmidt et al., 2013).

The processing and modelling of spatial (e.g. maps) and biological (e.g. biomolecular sequencing) large input data are highly demanding on computer and software's capabilities that only modern platforms can provide (e.g. High-Performance Computing).

[iv] *Vertical distribution patterns of soil properties in depth as defined by* (Pendleton and Jenny, 1945).

In order to frame the methodological procedures used throughout our investigation we describe an overview of (i) data modelling and mapping in the current context of the identification of microbial biogeographical patterns, (ii) modern biomolecular techniques and technologies used for microbial identification and (iii) metrics applied to estimate microbial diversity (e.g. Chao1) afterwards used as the input variables to explain observed/predicted on mapped soil microbial communities.

Aims

Under the frame of the discussion above, this investigation hypothesises that there is a co-spatial correlation between soil and microbial diversity gradients. On the basis of this hypothesis our general aim was:

To evaluate geographical patterns of soil microbial diversity across different agroecological zones of New South Wales (NSW)

Hence, the specific objectives of this thesis were:

1. The design and sampling of two orthogonal transects defining the study area including *disturbed* and *undisturbed* ecosystems.
2. Identification of *bacteria*, *fungi* and *archaea* communities using DNA sequencing approach and subsequent measurement of microbial α-diversity.
3. Measurements and predictive estimation of soil physicochemical attributes and other environmental covariates by using laboratory measures, NIR spectroscopy, and remote sensing data.
4. Assessment of observed soil microbial diversity linear, non-linear and multidimensional relations with soil physicochemical soil attributes along both the longitudinal and latitudinal environmental gradients of the study area.
5. Modeling, prediction, and mapping of *bacteria*, *fungi* and *archaea* α-diversity using *Scorpan model* across NSW.

MODELLING MICROBIAL BIOGEOGRAPHICAL PATTERNS

The close relations between soil and microbial communities have been thoroughly revised by many authors (e.g. Global Soil Biodiversity Atlas) (Orgiazzi et al., 2016). Therefore, there is a vast amount of evidence shedding light the closeness but also multivariate of this biotic-abiotic association. Within several references aimed to evaluate soil microbial geographical patterns at large scales (e.g. landscape, regional and continental scales), there is consistent evidence suggesting that soil microbial communities follow spatial patterns. Most commonly, these patterns have been found by evaluating microbial local diversity (richness, evenness, and phylogenetic diversity), relative abundances, taxonomic compositions or dissimilarities (species and divergence based) among communities.

Regarding the experimental designs, different scenarios have been evaluated following environmental gradients along latitudinal/longitudinal transects (Liu et al., 2008; Xia et al., 2016; Xiong et al., 2012), landscapes (Constancias et al., 2015b; de Vries et al., 2012; Oline et al., 2006), regions (Dequiedt et al., 2009; Wakelin et al., 2010), countries (Griffiths et al., 2011) and as far as some attempts of intercontinental scales (Tedersoo et al., 2014).

The evaluation of soil attributes driving microbial communities is usually done within the context of an environmental gradient (de Gannes et al., 2015; Dimitriu and Grayston, 2010; Richter et al., 2014; Wakelin et al., 2010; Xia et al., 2016) but also following specific gradients of

properties such as soil pH (Rousk et al., 2010; Xiong et al., 2012).

There are others studies that have included specific soil attribute gradients in their analysis such as hydrology (Krause et al., 2013), elevation (Zhang et al., 2013, 2015) or aridity (Maestre et al., 2015). In other experimental designs, land-uses/managements were a principal factor in evaluation (Constancias et al., 2015b; Kasel et al., 2008; Lauber et al., 2008; M.J. Swift M. van Noordwijk et al., 2004; Shahbazi et al., 2013).

From these studies, there is a general acceptance that the soil attributes found as main drivers controlling microbial patterns are *inter alia,* soil pH (Kuang et al., 2013; Lauber et al., 2008; Liu et al., 2014; Xia et al., 2016; Zhang et al., 2013, 2015), soil nitrogen (Xi and Bloor, 2016), soil organic carbon (Xi and Bloor, 2016), and environmental variables such us as temperature (Zhou et al., 2016) and elevation (Nottingham et al., 2016; Zhang et al., 2015)

Noteworthy, the results summarized above only shed light on the best-ranked attributes found and other less influential environmental or soil variables were found. Nevertheless, given these close relations between soil and microbial patterns it would be expected that microbial geographical patterns would resemble similarities with soil spatial patterns, i.e. both are co-spatially related.

Soil spatial patterns have been exhaustively studied for the last 130 years, hence there is a vast amount of work defining the distribution of soil properties in the space and the factors

controlling their predictions (Brevik et al., 2016; Dokuchaev, 1883; McBratney et al., 2003; Pendleton and Jenny, 1945; Trangmar et al., 1986; Webster et al., 1990).

In this regard, it is not unexpected that recent works in the area of microbial distribution have shown that they have consistent geographical patterns already observed in the areas of applied soil geo-statistics i.e., microbial spatial autocorrelation (Krause et al., 2013; Robeson et al., 2011). Consistently, and to the best of our knowledge, the latest studies that attempt to picture microbial geographical distribution have only made use of spatial approaches like *inverse distance weight and kriging* (Constancias et al., 2015b; Griffiths et al., 2011; Orgiazzi et al., 2016; Tedersoo et al., 2014).

McBratney et al. (2003) made a comprehensive revision of the different methodologies for representing the soil distribution in the landscape. In this review, McBratney et al. (2003) outline the benefits and limitations of purely spatial approaches such as *kriging techniques*. In their work, these authors also recommended a detailed empirical modeling and mapping procedure best known as the *Scorpan model approach* inspired by the concept of the soil formation factors coined by Hans Jenny (Pendleton and Jenny, 1945). The advantages of this approach over geostatistical approaches are that it considers the soil distributions in an entirely quantitative way to other spatially referenced factors as follows:

$$S = f(S, C, O, R, P, A, N)$$

where,

- **S**: soil, other properties of the soil at a point
- **C**: climate, climatic properties of the environment at a point
- **O**: organisms, vegetation or fauna or human activity
- **R**: topography, landscape attributes
- **P**: parent material, lithology
- **A**: age, the time factor
- **N**: space, spatial position

If we consider the well-documented relation between soil microbial and physicochemical gradients we can hypothesize that microbial communities will have a shared influenced by the before mentioned factors. Consequently, one of the objectives of this investigation will be to map the microbial diversity across NSW using the *Scorpan* approach as detailed in Chapter 4.

OVERVIEW OF BIOMOLECULAR SEQUENCING APPROACHES

Early microbiology: from microscopy to pure cultures

The combination of soil DNA extraction followed by the sequencing characterization of the small subunit ribosomal DNA (SSU rDNA) genes is likely the most widely used of the *culture-independent methods (CIMs)* available for microbial taxonomic identification. In recent years, the combination of these methods has allowed large-scale exploration on microbial communities of any environment (e.g. aquatic, soil or air ecosystems). Despite some pending bias corrections when working with environmental samples (e.g. referenced criteria about DNA quality), DNA sequencing analysis allows not only the production of high-resolution data (e.g. thousands of DNA sequences) but also uniform information for global analysis (Gasc et al., 2015; Green and Keller, 2006; Hazen et al., 2013; Nesme et al., 2016; Rastogi and Sani, 2011; Zarraonaindia et al., 2013).

Regardless of the necessary improvements behind genomics-based applications, there is no doubt that the introduction of biomolecular techniques by the early 1980's launched a new era in every research area exploring microbial communities but even more essentially for those areas exploring complex ecosystems such as soil environments (Simon and Daniel, 2011; Torsvik, 1980). Prior to biomolecular techniques, all the accumulated knowledge for 300 years of early

microbiology - including the first study of soil microorganisms - were developed based on both microscopy and pure cultures on artificial media. Microscopy was used since the first fungus and bacteria were detected under the microscope by Robert Hook in 1665 and Antoni van Leeuwenhoek in 1675, respectively (Gest, 2004). Two hundred years later, Franz Unger introduced the use of pure cultures on artificial media which was later enhanced and formalized by Robert Koch's in the 1880s by investigating diseases causality (Handelsman, 2004). Moreover, the incorporation of this bacteriological technique for laboratory microbiology demonstrated that known and unknown microbes resisted being cultured and since then microorganisms have been divided between cultured and uncultured ones (Handelsman, 2004).

Modern microbiology: from culture-dependent to culture-independent methods

The first steps in modern microbiology extended our understanding of cultured microorganisms – e.g. studying DNA from single cells. But afterward a fundamental contribution was to open access to uncultured microbes. Hence, modern molecular studies started an important differentiation between the analysis of microbes by the traditional 'culture-dependent' method (direct cultivation) and the novel culture-independent alternatives *via* biomolecular manipulation (Daniel, 2005; Dokic et al., 2010; Nesme et al., 2016; Torsvik L., 2002). An equally important distinction of the modern molecular era was the incorporation of technological advances such as high-throughput sequencing (HTS) platforms and bioinformatics tools. All these advances together allowed for the first time the analysis of microbial communities at the

resolution of environmental DNA (e.g. soil DNA) (Zhou et al., 2015).

Culture- dependent methods

These protocols rely on the isolation of the microorganisms into an artificial nutrient-rich or nutrient-poor media in laboratory conditions (Daniel, 2005; Zarraonaindia et al., 2013). For over 120 years, the artificial growth medium has been agar Petri plates (Tanaka et al., 2014). A pure culture of individual cells is created on this agar surface, and later they are easily separated for their growth, division, and colonization by forming thousands of clones. These clones are then quantified using the traditional counting method "the colony-forming unit" (CFU) (Joseph et al., 2003).

Culture- independent methods (CIMs)

CIMs are basically founded on the direct isolation of biomolecules (e.g. nucleic acids, proteins, lipids, etc.) from either an individual genome (e.g. cell, organism) or a metagenome - a collective of genomes derived from any microbiome or environment (e.g. gut, lake, soil, etc.) (Simon and Daniel, 2011; Torsvik L., 2002). Once isolated the genomics products, what is following in downstream procedures will largely depend on the purpose of the investigation on the target microbial community (e.g. taxonomy, phylogenetic analyses, functional diversity, etc.). For example, most of the microbial taxonomic and functional profiling have relied on the DNA sequencing of the SSU rDNA gene resulting vastly used for diversity estimations in both cultured and uncultured methods (Hazen et al., 2013; Mendoza et al., 2014; Zarraonaindia et al., 2013). Nucleic acid, DNA, and RNA have been the biomolecule most largely extracted to perform

microbial analyses so far. However, depending on the nature of the exploratory analysis also others biomolecules such as protein, lipids, metabolites play an important role and they are very useful these days (Zhou et al., 2015).

A more comprehensive insight into microbial exploration could require a combination of *CIMs* or even their application in conjunction with culture-dependent strategies. Both culture-dependent and culture-independent strategies might be complementary in particular cases such as when linking microbial activities to genes or to metabolically active enzymes (Blagodatskaya and Kuzyakov, 2013; Green and Keller, 2006; Joseph et al., 2003).

Large-input microbial analyses: CIMs, biomolecular sequencing, and bioinformatics

Modern advances in microbial studies do not solely rely on molecular laboratory techniques but also on biomolecular sequencing technologies and bioinformatics. CIMs involves innovations in the laboratory strategies, e.g. PCR amplification-based, gene cloning, sequencing of 16S rRNA genes and denaturing gradient gel electrophoresis; biomolecular sequencing technology involves a range of high-throughput sequencing (HTS) platforms where molecular information is decoded; and bioinformatics tools comprise the computational capacities (e.g. software) to translate and analyse the biomolecular codes (e.g. BLAST, QIIME) (Gasc et al., 2015; Henry et al., 2011; Simon and Daniel, 2011; Xu et al., 2014; Zhou et al., 2015).

This advanced trilogy has revolutionized multidisciplinary areas of microbial studies and even opened new ones (e.g. metagenomics, proteomics) (Schneider and Riedel, 2010; Simon and Daniel, 2011; Zhou et al., 2015). Environmental microbiology has been especially benefitted with modern advances since the isolation of microorganisms from natural environments is one of the main challenges, particularly, from soil samples (Lombard et al., 2011; Zhou et al., 2015). Indeed, only after ten years using CIMs was valued the wide scope of the soil genetic diversity. Since then not in vain the uncultivated majority of soil microbes is denoted as the 'microbial dark matter' (Dokic et al., 2010; Gasc et al., 2015; Green and Keller, 2006; Larsen et al., 2012; Lombard et al., 2011; Torsvik L., 2002).

It is estimated that the proportion of microbial diversity directly cultivated with standard techniques is less than 5% of the total present in the biosphere (Zarraonaindia et al., 2013). More specifically in the case of soil ecosystems, there are references showing that bacterial communities growing on agar media can vary between 0.1 to 1% (e.g. pristine forest soils) to up to 10% (e.g. arable soils) of the total diversity (Dokic et al., 2010; Torsvik et al., 1998). The most popular reference documented in the literature referencing this methodological limitation states that more than 99% of prokaryotes in the environment cannot be cultured in the laboratory (Gasc et al., 2015; Green and Keller, 2006; Nesme et al., 2016; Schloss and Handelsman, 2005). This phenomenon is often referred as the "great plate count anomaly" (Ayrapetyan and Oliver, 2016; Tanaka et al., 2014). Therefore, the uncultivated majority of soil microbes is just being uncovered (especially rare communities) by genetic techniques over the last 35 years, (Dokic et al., 2010).

The first direct extraction and purification of DNA from a soil environment are attributed to Vigdis Torsvik with the isolation of bacterial DNA in 1980. Since then, our narrow understanding was revealed but at the same time emerged a systematic and comprehensive vision of microbial communities as part of a major 'network system' (e.g. biogeochemical processes, soil food web, etc.). Therefore, an integral understanding of their multifunctional biotic –abiotic interactions in the context of their environments (i.e. metabolism, physiology, ecology, whole-genomes, genetic diversity, functional diversity, evolution, among others) has been needed (Gasc et al., 2015; Loman and Pallen, 2015).

To obtain 'systematic' and 'comprehensive' knowledge about microbial environments, modern studies evolved to explore at wider scales. Even the scientific perspective evolved to more specific research fields such as the so-called 'omics' sciences, e.g. genomics, transcriptomic, lipidomics. Indeed, the suffix 'omics' here conveys the notion of systematic and a comprehensive study (Akondi and Lakshmi, 2013; Gugerli et al., 2013; Nesme et al., 2016; Zarraonaindia et al., 2013; Zhou et al., 2015). Moreover, this disciplines when extended to analyze organisms collectively (i.e. working at a meta-scale) used 'meta' as connotation, e.g. soil metagenomics (Simon and Daniel, 2011).

Towards this direction, recent studies have attempted to increase their experimental resolution (e.g. larger spatial and temporal scales gradients understanding biogeographical distributions) and to increase biomolecular techniques efforts, e.g. deepness in DNA sequencing analyses for microbial identification (Gugerli et al., 2013; Sinclair et al., 2015). For example, since circa 1998

soil metagenomics became a discipline which aimed to identify the total biological entities within a complex soil sample (Nesme et al., 2016; Zarraonaindia et al., 2013). Nowadays there is more interest in large-scale surveys focused on identifying microbial communities based on their genetic material (e.g. DNA, RNA, etc.) but as part of a collective soil genome. Daniel (2005), explained that "*theoretically, the microbial DNA isolated from a soil sample represents the collective DNA of all the indigenous soil microorganisms*', the so-called *soil metagenome*.

Technological improvements had largely contributed to make possible a large input microbial research. This has been especially important not only for microbial identification based on biomolecular sequencing and corresponding decoding but also for the management of the big amount of data, e.g. clone libraries (Reuter et al., 2015; Simon and Daniel, 2011; Zhou et al., 2015).

In the course of the last forty years, all the major milestones and directions coursed in microbial investigations remained determined to technological improvements on sequencing platforms – so called High Throughput Sequencing technology (HTS) (Gasc et al., 2015; Loman and Pallen, 2015). The first sequencer was developed by Frederick Sanger in the 1970s and was the most widely used for 25 years (Shokralla et al., 2012). The latest innovations reached during the last 20 years, however, were the ones that delineated the most representative features of the commercial sequencing platforms available these days. Loman and Pallen (2015) recently published a detailed timeline overview on these technologies applied for bacterial genomes sequencing analyses. By using Figure 1-1 below, these authors described the three main

technological revolutions that transcended to the way how microorganisms are studied at a meta-scale, i.e.:

- The First Revolution: whole-genome shotgun sequencing platforms, e.g. Sanger Shotgun Sequencing instrument.
- The Second Revolution: high-throughput sequencing technologies also known as next-generation sequencing platforms, e.g. 454 Pyrosequencing and Illumina Sequencing instruments.
- The Third Revolution: single –molecule sequencing platforms, e.g. Oxford Nanopore Sequencing instruments.

As the principles of sequencing pipelines evolved, each of these technological strategies developed crucial innovation on their technical capacities (e.g. efficiency, rapidness, the length of readings, etc.; Figure 1-2). Therefore, there are different formats varying between instruments (e.g. 454 sequencing vs Illumina Sequencing) as well as these formats are variable among different platforms in harmony to their in deep sequencing capability (MiSeq platform and HiSeq platform on Illumina Sequencing) (Loman and Pallen, 2015; Zarraonaindia et al., 2013; Zhou et al., 2015).

Reuter et al., (2015) have recently reviewed the usage of HTS technologies according to their cost, performances, and capacities. As they illustrated in Figure 1-2, there is an important diversification of their capabilities during the last decades mostly referred to the technical features such as outputs per run (megabytes produced). For example, we can see different

options under Illumina technology (e.g. HiSeq, MySeq, etc.). For example, MySeq platform, the one used for our sequencing performances, generates sequences of no more than 300 base pairs in length and signifies outputs near to 10,000 Mb.

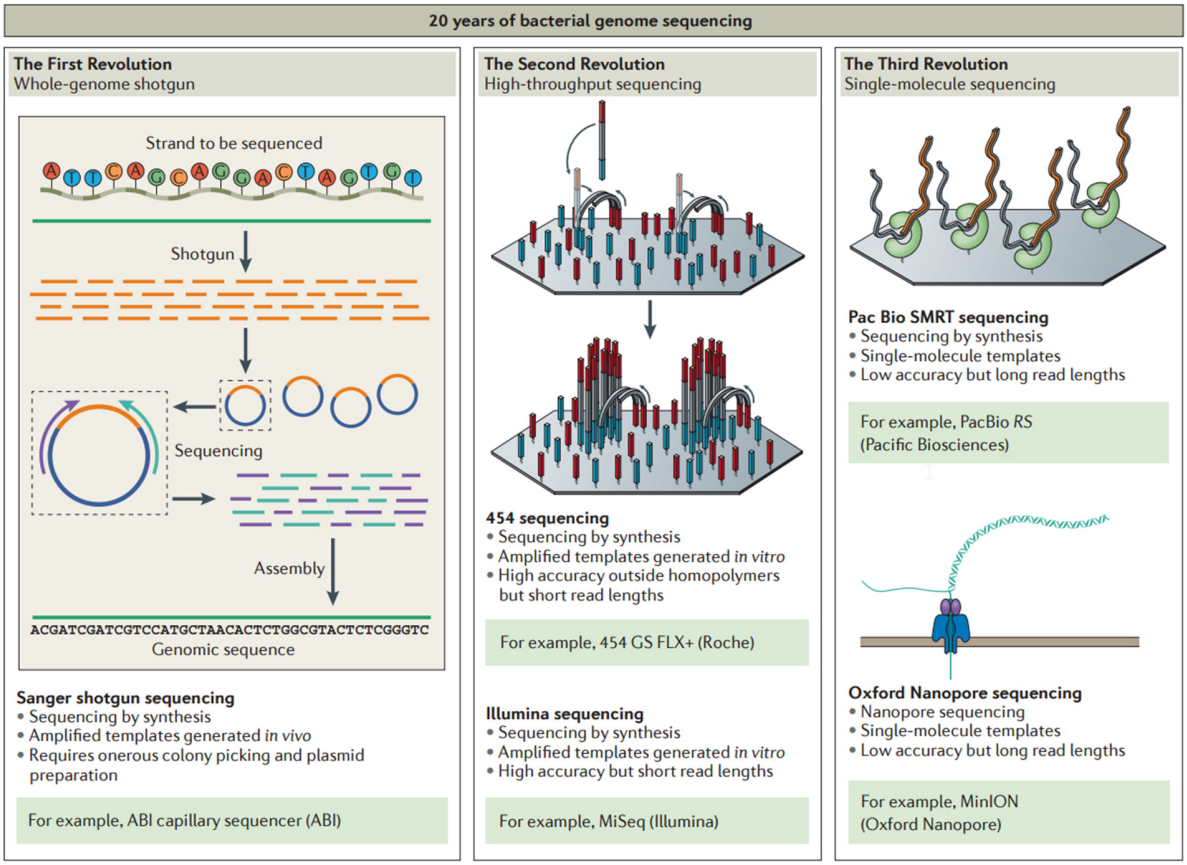


Figure 1-1. The first two decades of bacterial genomics analyses schematized by Loman and Pallen, 2015. The three revolutions in sequencing technology that have transformed the landscape of bacterial genome sequencing are as follows: Whole-genome shotgun, High-throughput sequencing and, Single-molecule sequencing.

The features of these sequencing platforms, their commercial cost per sequence, and the computing performance capacities required for the subsequent data analysis became an

important consideration when performing meta-analyses on microbial studies, especially for metagenomics science.

As indicated earlier, the use of these instruments has largely transcended the way microorganisms are studied today at a meta-scale. Indeed, the biomolecular sample to be used by these HTS technologies has changed their protocols for preparation. These protocols are used by specific meta-omics disciplines (e.g. metaproteomics) but most generally are applied to reconstruct microbial metagenomes and evaluate the structural and functional organisms composing different microbiomes (Gasc et al., 2015) - such as the case in this research.

There are different workflows for laboratory sample preparation and sequencing instrumentation. Among these alternatives perhaps the most commonly used are *metagenomics-based*, *metabarcoding*, *culture-based, single-cell genomics and gene capture approach*. Strengths and weaknesses of them are briefly compared by Gasc et al., 2015 whose diagram is shown in Figure 1-3.

All the alternatives above can be used to increase our knowledge and contribute to the soil metagenomics' discipline - i.e. identify the biological entities within a soil metagenome. In this respect, owing to multiples misunderstandings we contribute to clarified that must be distinguishable the term of 'metagenomics' when it refers to the whole discipline (e.g. soil metagenomics) and when it makes reference to the laboratory technique (*metagenomics-based approaches)*. This conceptual and methodological clarification has been discussed in detail by

Mendoza et al., 2014 and Taberlet et al., 2012 as well as in others instances (e.g. panel discussion at the 2nd Thünen Symposium on soil metagenomics, Braunschweig, Germany, 2013).

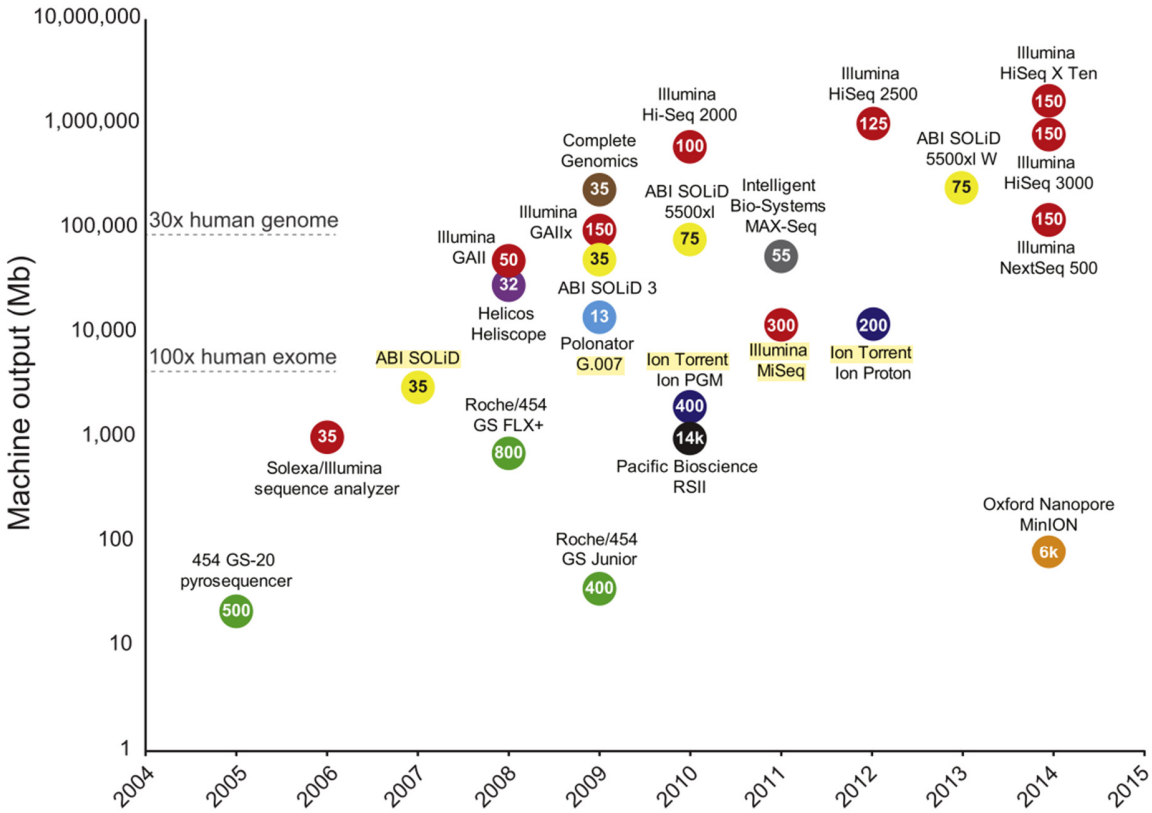


Figure 1−2 Timeline and comparison of Commercial HTS instrument published by Reuter et al., 2015. HTS instruments release dates *versus* machines output per run. Numbers inside data points denote current read lengths. Colour coded shows different sequencing platform available these days.

It is argued that due to the two approaches most widely used for DNA characterization is that metagenomics 'discipline' has derived in two fields: one based on DNA metabarcoding approaches and another one based on metagenomics-based approaches (Mendoza et al., 2014).

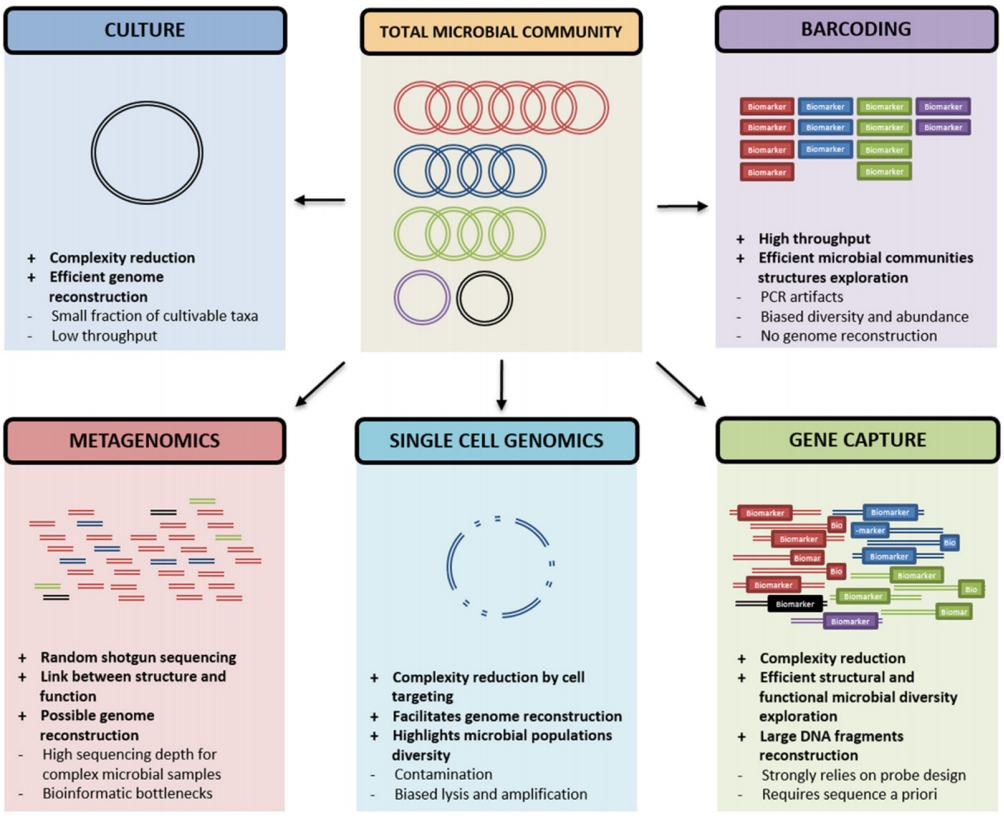


Figure 1–3. Strengths and weaknesses of various molecular approach to link identity and function in metagenomics samples (Gasc et al., 2015).

DNA metabarcoding approach

The principle is based on the sequencing of *'a priori'* defined DNA region of the soil metagenome, i.e. target DNA genes or regions. For example, those DNA regions containing the 16S rDNA and/or ITS genes. Technically, these target genes or regions can be sequenced either (i) directly on the soil metagenome or (ii) indirectly by sequencing 'clones' of them obtained from the soil metagenome. (i) Direct DNA sequencing works on the basis of shotgun sequencing technology, in which, the target DNA regions are randomly selected. (ii) Indirect DNA sequencing implies: (a) to barcode the target DNA region with molecular markers and (b) amplify this target to obtain

its clones *via* PCR-based methods, i.e. 'amplicons sequencing strategy' (the strategy applied in our research). Then, these amplicons are sequenced by HTS platform (e.g. Illumina) (Gasc et al., 2015; Larsen et al., 2012; Mendoza et al., 2014; Schmidt et al., 2013; Shokralla et al., 2012; Zarraonaindia et al., 2013).

Metagenomics-based approach

In this case, the principle is based on a direct sequencing of the soil metagenome using shotgun sequencing platforms. Mendoza et al., (2014) have indicated that a fundamental difference between 'metagenomics' and 'DNA metabarcoding' approaches is the data generated. The former provides additional genomic-scale information enables not only taxonomic identification but also the functional characterization of the environmental sample. For example, Simon and Daniel, (2011) indicated how the construction and screening of metagenomics libraries have resulted in the identification of many novel biocatalysts, including lipases/esterases, cellulases, DNA polymerases, proteases, and antibiotics.

Finally, these protocols for DNA's sequencing sample preparation (among others) can be more or less recommended depending on both our research purposes and the advantages and disadvantages of HTS platforms for such purposes. Certainly, the use of DNA-metabarcoding approaches is recommended for biodiversity and community structural studies, whereas, metagenomics-based approach are more appropriated when linking functions to structural diversity (e.g. this study) or whole-genome reconstruction (Figure 1-3). Moreover, the DNA-metabarcoding approach in combination with Illumina HTS platform has been highly advised for

taxonomic studies in large-scale studies. It is argued that in this way a high number of sequencing reads is provided but shorter in length which is considerably less time-consuming for bioinformatics processing (Schmidt et al., 2013; Zarraonaindia et al., 2013).

In sum, the Illumina DNA metabarcoding approach based on amplicon sequencing (PCR-based) has been the strategy used in our microbial identification analysis. In general, sample preparation for DNA sequencing requires three main steps; (1) template preparation, (2) sequencing on the platform and, (3) interpretation of the biomolecular information. All these specifications are presented in Chapter 3 in the methods section.

OVERVIEW OF SOIL MICROBIAL DIVERSITY METRICS

Introducing microbial diversity analyses

In this research, microbial diversity has been estimated to assess the structure of the soil microbial communities across different habitats in NSW. The structural diversity of the soil samples has been characterized using *quantitative* and *qualitative* measures to inform about species richness (incidence-based, i.e. presence-absence), evenness (distribution of abundances) and phylogenetic diversity among the microbial communities as also how their distribution change across geographical distances and environmental gradients.

There are several variations on how to characterize these diversity patterns on biological communities (e.g. rank-abundance curves, indices) (Lozupone and Knight, 2008; Magurran, 2004; Morris et al., 2014; Nemergut et al., 2013; Tuomisto, 2010). Biodiversity, a multidimensional property of natural systems, is *qualified* and *quantified* using diversity estimators (e.g. indices, coefficient, plots, etc.). These biodiversity measures indicate how rich and even a given community is and how similar and/or dissimilar two or more communities are when compared. Indeed, thanks to a collective and multidisciplinary effort to characterise 'diversity' we count today with a set of 'biodiversity indices' able to enlighten about diversity patterns from different aspects of interest (e.g. abundance, dominance, phylogenetic-relatedness, commonness, rarity, etc.) and different spatial and temporal scales (e.g.

agroecological zones) (Escarguel et al., 2011; Magurran, 2004; Morris et al., 2014; Whittaker, 1972).

By itself there is no single 'diversity metric' (parametric or non-parametric) flexible enough to qualify and quantify the entire extent of the diversity concept (Magurran, 2004; Morris et al., 2014). In this regard, many authors advocate differentiating the use of 'diversity' in the entire context of its definition and a 'diversity index' as the metric use to estimate the first one. Indeed, it is required the use of different indices to cover a deeper characterisation of the entire diversity and only one 'diversity index or metric' may not be sufficiently informative for such purpose (Tuomisto, 2010). Reasonably, the origin of each of these 'diversity indices' was not motivated from biological and ecological explorations but from other research areas and disciplines – e.g. one of the most common diversity measures, Shannon index, was developed to estimate the uncertainty (entropy) in telecommunication messages.

According to their different nature, each of these diversity indices carries their own strengths, weaknesses, and perspective on what is actually defining a greater or lesser diversity in a given community. Each of these estimators has its own principles and statistics, but all of them accomplish three main assumptions that must be applied for biodiversity measurement: (i) all species (OTUs) are equal (ii) all individuals are equal (DNA sequences) and (iii) comparable unit of measure for abundance data (e.g. only DNA sequences or only biomass data) (Lozupone and Knight, 2008; Magurran, 2004). On the other hand, the main differentiation among these estimators is the extent to which they weigh the 'richness' and 'evenness' aspects of diversity

(Lozupone and Knight, 2008; Magurran, 2004).

Despite being a multidisciplinary task, finding the most suitable 'biodiversity metrics' has become considerably more challenging for microbial ecology during recent years. The insertion of new genetic techniques has opened new dimensions counting the uncountable diversity of microorganisms inhabiting highly heterogeneous habitats such as natural environment and, particularly, soils (Hughes et al., 2001; Lozupone and Knight, 2008). Consequently, the new format evaluating diversity patterns on the genetic information of miscellaneous and tiny organisms has meant to add new considerations to able these analyses. For instance, some of the new matters have been *(i) to define the basic unit for diversity measurements* (e.g. *cluster of sequences instead of species), (ii) to extend diversity analyses to genetic variations* (e.g. *divergence lineage between taxa), (iii) to calibrate the sampling effort to properly represent a given community* (e.g. *number of sequences required to copiously represent a given community), (iv) to define the criteria of comparison between communities* (e.g. *similarities and dissimilarities), among others* (Hughes et al., 2001; Morris et al., 2014).

Transversally, each of these components affects directly or indirectly the existence, applicability, and interpretation of the diversity estimators, e.g. indices/coefficients, plots or curves. By incorporating new formats for diversity analysis and taking advantage of the massive sequencing data derived from large-scale microbial surveys, these different metrics have evolved to cover more complex perspectives in biodiversity descriptions.

As biodiversity metrics we changed over the time, their *pros* and *cons* have been critically reviewed by different authors (Magurran, 2004; Morris et al., 2014). A comprehensive and complete analysis was made in 2004 by Magurran. In this book, Magurran reviewed all aspects of measuring biodiversity: origins, principles, models, surrogates, assumptions, concepts and important applications of the most popular biodiversity indices. A great part of such references is used in this manuscript. However, Magurran (2004) emphasized that she did not review measurements applied to microbial diversity analyses based on molecular techniques and phylogenetic variations. Other authors have explored these more contemporaneous indices and estimators (Hill et al., 2003; Hughes et al., 2001; Lozupone et al., 2007; Lozupone and Knight, 2008; McMurdie and Holmes, 2014; Morris et al., 2014).

A complete and well-structured review on diversity measures focused on microbial communities was published by Lozupone and Knight (2008). In fact, since 2007 Lozupone et al. have proposed to organize all the diversity indices variations as shown in Table 1-1, in which diversity measures are framed into three main distinctions whether diversity is: (i) analysed in terms of *species-based measures*, by considering all taxa as equally related and excluding distance relatedness among them from the analysis, and/or *divergence-based measures*, by quantifying into the analysis the distance among all taxa as a diversity component (ii) measured qualitatively, only based on presence-absence data, and/or quantitatively, including frequency-abundance data, and (iii) analysed within a given community as the α-diversity and/or among different communities as the β-diversity. More about features, parallels, contrasts and extend beyond diversity metrics is described below.

Table 1−1. Categories of diversity measurements as described by Lozupone and Knight (2010).

	Measurement of diversity within a single community(αdiversity)	Measurement of diversity shared among communities(β diversity)
Only presence/absence of taxa considered	Qualitative α diversity (Richness) Species-based: Chao 1, ACE, Rarefaction Divergence-based: Phylogenetic Diversity (PD)	Qualitative β diversity Species-based: Sörensen index Jaccard index Divergence-based: Unweighted UniFrac Taxonomic Similarity (Δ_S)
Additionally, accounts for the number of times that each taxon was observed	Quantitative α diversity (Richness and/or Evenness) Species-based: Shannon's index Simpson's index Divergence-based: Theta	Quantitative β diversity Species-based: Sörensen quantitative index Morisita-Horn measure Divergence-based: Weighted UniFrac F_{ST} DPCoA

Diversity measures: from species-based to OTU-based methods

The first distinction in modern microbial analyses is the fact of being measuring diversity on the basis of 'genetic sequences' instead of 'species' itself. Historically, microbial diversity has been characterized by species-based methods, i.e. those using species as the basic unit of measure (Lozupone and Knight, 2008) and/or, others indirect ones, e.g. biomarkers methods such as *Phospholipid Fatty Acid Analysis* (PLFA). The quantification of diversity by species-based methods has been generally defined in terms of presence/absence (richness) and frequency-abundance (evenness) of the individuals living in a given sample. Universally, the indices used in these analyses are Shannon or Shannon-Wiener (Shannon, 1948), Chao1 (Chao, 1984), Simpson

(Simpson, 1949) and few others surrogates of them (e.g. Simpson's dominance index) (Morris et al., 2014).

Species-based methods have been the traditional scheme used for diversity estimation in circumstances when microorganisms were mostly identified by culture-dependent methods and the microbial species were differentiated phenotypically and/or by hybridizing their DNA to replicate the same species to truly classify the one it was. However, these analyses increased its complexity when we began working with thousands of microbial DNA sequences at once or microbial molecular fingerprinting patterns for their diversity characterisation. Moreover, this kind of genetic information has opened new edges on which rely diversity assessments such as the overall 'relatedness among genomes' (Lozupone and Knight, 2008).

Species-based diversity measurement based on genomes-relatedness on DNA sequences has been particularly advantageous in prokaryotes whose primary reproduction form is generally asexual. This type of reproduction able *bacteria* and *archaea* to recombine genes of very distant species using horizontal genes transfer (HGT) which has complicated their phenotypic differentiation and so diversity characterisation when working on the basis of culture-dependent methods - apart from the fact that it can be unviable for most of the species as explained in earlier in this Chapter. In contrast, the analysis of microbial diversity on the relativeness of their genomes greatly solved this particular issue with prokaryote, as well as, provide more precision when classifying microbes through the assemblage of their DNA sequences.

On its own, this new format for searching into microbial genomes has introduced additional concerns for diversity measures. One of the ongoing discussion about counting and classifying microorganisms based on 'genetic sequences' is what defines a species (Gevers et al., 2005). In fact, this definition is still being debated since the boundaries for a given DNA sequence of whether an organism belongs to one or another species is not obviously delineated (Konstantinidis et al., 2006; Tuomisto, 2010). This arrangement is typically made by defining a similarity threshold (e.g. 97% as minimum to equal an empirical limit for same species when isolated by culturing methods) by which are clustering similar sequences within a determined species (Gevers et al., 2005; Lozupone, 2007; Martin, 2002; Mendoza et al., 2014). This has led to questioning whether the number of 'species' is truly represented and therefore the concept of the operational taxonomic unit (OTU) is preferable as the basic unit for measuring diversity instead of 'species' itself at this taxonomic level. OTU can be any of the basic units of diversity measurements depending on the methodology applied to study the microbial diversity. For example, an OTU can either be representing the number of DNA sequence similarity groups or the number of unique terminal restriction fragments (when microbial communities are profiled using fingerprinting techniques) (Hughes et al., 2001).

Diversity measures by genetic variation: *species-based vs divergence-based* methods

Sucha as the microbial identification-classification has evolved also the diversity measurements methods have been expanded to cover new exploratory analyses such as genetic divergence

between taxa (Martin, 2002). As introduced above, species-based diversity methods (i.e. Shannon, Chao1, Simpson, etc.) on the basis of DNA sequences have essentially assumed that the genomes-relatedness is equal among the different species, i.e. it does not consider into account the genetic divergence between the different microorganisms. In other words, these methods quantify different taxa but not the degree of distance between the lineages of these taxa. This consideration led to developing divergence-based methods for diversity characterisation. By means of these metrics is given higher diversity values to communities that harbor more distant lineages, e.g. those which are more phylogenetically diverse (Lozupone and Knight, 2008).

More recently, divergence-based methods have been presented as a powerful tool for diversity characterisation. As explained above, sequence similarity (which can be cut-off at different threshold values) often correlates positively with phenotype but genetically microbes from a given community can differ enormously. This is a valuable information which can be a turning point on diversity characterisations (Martin, 2002). This genetic divergence diversity can be analysed at three levels: (i) sequence distance by measuring the separation between two sequences in terms of the number of nucleotides, (ii) phylogenetic distance by measuring the separation between two taxa in terms of the amount of branch length in the phylogenetic tree; and (iii) taxonomic distance by measuring the number of edges that separates two taxa in taxonomic tree (Lozupone, 2007). Diversity measures based on phylogenetic distance analysis are more broadly applied and, in this cases, the Phylogenetic Diversity (PD) index introduced by Faith in 1992 seems to be the most frequently applied (Cadotte et al., 2010).

Diversity measures for richness and evenness characterisation: *qualitative vs quantitative* methods

Another important variation among diversity indices is that these measures can be either weighted or unweighted by the relative abundance of the individuals and so, respectively, characterise richness (only types of individuals) or evenness (abundance distribution of the individuals) patterns of the community (Cadotte et al., 2010; Lozupone, 2007). As reviewed by Lozupone (2007), there are numerous biological studies with meaningful differences in diversity results when counting or not the relative abundance of the individuals. This author referred both types of metrics as *qualitative measures;* when diversity is estimated only based on species-based incidence, i.e. presence/absence; and *quantitative measures* when the diversity estimation is weighted by the relative abundance of each taxon. *Qualitative measures* compute 'richness of the community' whereas *quantitative measures* can evaluate the 'evenness of the community' and, therefore, in other words, heterogeneity, dominance (obverse of evenness), commonness and rarity (Lozupone, 2007; Lozupone et al., 2007; Magurran, 2004). In this thesis, we discuss diversity indices using the organization and terminology proposed by Lozupone (2007).

Abundance defining commonness and rarity

Weighting by the relative abundance of the individuals, i.e. on quantitative measures, has been an important aspect of the way how microbial diversity is studied nowadays – which is largely dependent on the sampling effort and data collection subjected to time and cost-effectiveness.

It is broadly accepted that within biological communities, the abundance of different species are significantly variable typically following a hollow-curve distribution with only a few abundant species and many rare species represented by the long tail of this curve (Figure 1-4) (Cadotte et al., 2010; Hughes et al., 2001; Lynch and Neufeld, 2015; McGill et al., 2007; Pedrós-Alió, 2012). This kind of illustration, 'so-called rank-abundance curve', is one of the most useful to depict richness but most particularly evenness based on the distributions of species abundances (Magurran, 2004).

This frequency distribution pattern is an important directive for modern microbial studies. Some authors have argued that the relative abundance of the species is the only factor that determines their importance when measuring diversity since it quantifies the effective number of types rather than the actual number of types (richness based on presence/absence). For these authors, 'diversity or biodiversity' are not actually measured if the proportional abundances are not taken into account and this kind of measures should be referred by other names, e.g. species richness, species turnover (Magurran, 2004; Tuomisto, 2010). In practice, excepting some particular cases, most of the microbial diversity investigations discard low-abundance taxa from their analyses (Lynch and Neufeld, 2015). For instance, low-abundance bacteria have been almost invisible for so many studies in which it is argued that a higher sampling effort would be required for considering them as important contributors for the diversity analyses (Pedrós-Alió, 2012; Sogin et al., 2006).

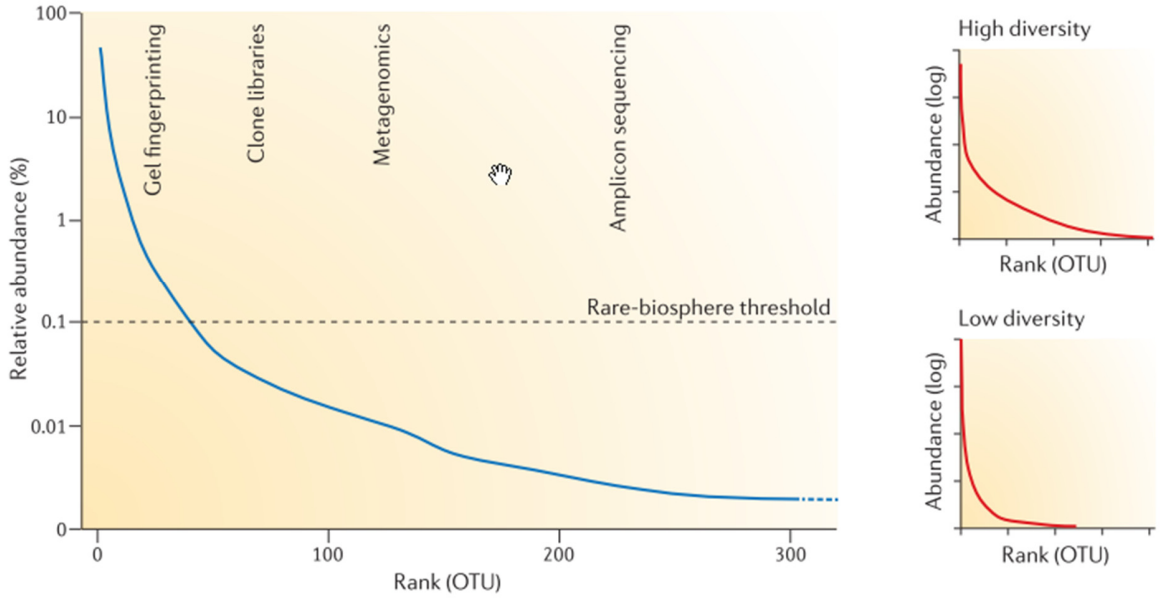


Figure 1-4. Rank-abundance curves. The graph on the left shows a schematic rank - abundance curve, which indicates the coverage depth ranking of different experimental techniques used in taxonomic surveys of microorganisms. The approximate threshold for the detection of rare-biosphere organisms is indicated. The graphs on the right show typical rank - abundance curves for high-diversity environments (such as soil) and low-diversity environments (such as feces), demonstrating the differences in the size of the long 'tail' that corresponds to the rare biosphere. OTU, operational taxonomic unit. Published by (Lynch and Neufeld, 2015).

On the contrary, Gaston (1994) developed a profound review about the rarity concept, emphasizing the importance of low-abundance individuals to evaluate endemicity, local population size, habitat specialization, etc. (Magurran, 2004). More recently, as result of modern microbial studies revealing higher resolution in the data, Sogin et al. in 2006 introduce the important role of the 'rare biosphere'. Since then, various authors have rescued the ecological contribution of these low-abundance taxa and unlimited source of genetic reservoir and functional diversity (e.g. major contribution to biodiversity and ecosystem resilience) that they signify for natural environments such as marine (Fuhrman, 2009; Pedrós-Alió, 2012) and soils ecosystems (Elshahed et al., 2008; Lynch and Neufeld, 2015; Reid et al., 2011). By 2009, the

American Academy of Microbiology has convened to begin the focusing of future work on these taxa (Reid et al., 2011).

Even though large-scale surveys advocated to explore the 'rare biosphere' are quite contemporaneous, there are some specific references to be considered when analyzing diversity in this kind of studies. For instance, more recent investigation in this area conventionally outlined the rare taxa between 0.1 and 1% of the total sequences count in the dataset (Figure 1-4) (Elshahed et al., 2008; Lynch and Neufeld, 2015; Pedrós-Alió, 2012). By 2004, Magurran had referenced that 'the rare species correspond to those that fall in the lower quartile of the species abundance distribution' according to Gaston's definition - which focuses attention on macroorganims. Furthermore, Lynch and Neufeld (2015) stated that changes in the abundance of dominant species/OTUs, in fact, can obscure our understanding of rare taxa dynamics arguing that they could behave cyclically (e.g. permanently-rare, occasionally-rare, transiently-rare, etc.). Thus, a determined community can turn to be abundant because of the surrounded conditions changed. In this cases, *qualitative diversity measures* can be more informative than *quantitative measures* about certain information as, for example, identifying diversity patterns of endemic communities.

Lozupone, (2007) indicated that either *quantitative* or *qualitative* diversity measures can lead to equally illuminating assumptions about the main factors structuring microbial diversity but with totally different results. For instance, when comparing communities, *quantitative measures* can elucidate changes in pattern abundances due to environmental changes (e.g. nutrient

limitation), whereas *qualitative measures* point out what can actually live in a certain environment (e.g. saline soil). Importantly, both *qualitative* and *quantitative* estimators for richness and evenness characterisations can be at the same time, weighted and unweighted by the genetic variation component using divergence-based methods as explained in the heading above.

In this research, the assessments of the rare taxa can provide useful information considering that the microbial dataset has been constructed on the basis of PCR amplicon sequencing in combination with NGS technology. Such combination is the best ranked of the molecular techniques to sample rare-taxa at a high resolution (Lynch and Neufeld, 2015; Pedrós-Alió, 2012). Additionally, as suggested by different authors, for a comprehensive understanding of microbial patterns we have applied both *qualitative* and *quantitative* measurements to our microbial sequencing datasets.

(i) Diversity measures adjusted by sampling effort

Sampling effort is another significant element when quantifying and qualifying diversity in particular when comparing diversities of different microbiomes. Diversity characterisation whatever the criteria of measurement is (richness or evenness with or without weighting on genetic relatedness) and the specific index applied (e.g. Shannon, Unweighted unifrac), the size of the sample (e.g. number of sequences found per soil sample) will affect the final results (Gotelli and Colwell, 2001; Lozupone, 2007; Magurran, 2004). One common understanding in statistical biodiversity is that the types of organisms observed increases with sampling effort until all types

(e.g. OTUs) are observed (Hughes et al., 2001; Magurran, 2004). In practice, this pattern is illustrated by plotting 'species accumulation curves', which on the basis of DNA sequencing data, record the cumulative number of OTUs as a function of the sequencing depth of sampling (e.g. from 1 to more than 10,000 sequences/sample as in the case of *bacteria* and *fungi* datasets in this survey). Thus, sampling effort and species accumulation curves are strictly associated.

The species accumulation curve (and others surrogate closely related, e.g. rarefaction curves, individual-based taxon sampling curves) provides useful information on the relationship diversity/sampling effort by showing the rate at which new species are found (Gotelli and Colwell, 2001; Hughes et al., 2001; Magurran, 2004). In general, species accumulation curves are constructed from left to right as the sampling effort increases in the *x*-axis. In general terms, these curves rise relatively rapidly at first and much more slowly in later samples as increasingly rare species are added when is expected to reach an asymptote (Gotelli and Colwell, 2001). The initial steepness of most accumulation curves reflects discovery of new taxa that correspond to more abundant organisms and therefore increased probability that they will be detected with minimal sampling. As the curve begins to plateau they detect new OTUs from lower abundance or more rare populations. The richer and more uneven the community, the longer it takes for the curve to level off, as new species continue to be found as sampling continues. For example, assuming a sufficient sampling effort, bacteria rarely approach the plateau but archaea communities can reach this level at the tenfold lower level of diversity than do those for bacteria (Reid et al., 2011).

Simultaneously, the curve can be 'normalized' by randomizing subsamples when is created, *with* or wit*h*out replacements, as new species are added (*without* replacement is recommended although penalize variance calculation) (Colwell, 2013). This procedure is suggested since the shape of the sampling curve is very sensitive to the order at which the subsamples are taken at each depth of sampling (e.g. 10; 50; 1,000; 10,000 sequences). For instance, a first subsample taken at 10 sequences depth can release independently a higher or lower number of observed species than a second subsample from the same dataset. Subsequently, the subsampling intensity would set up completely different curves leading to completely different interpretations of diversity patterns. For this reason, the accumulation curve is normalized by randomizing this subsampling protocol prior use of diversity estimators and indices – this procedure is completely different to randomization and rarefaction for comparative analyses ('rarefying') which is discussed below (Colwell, 2013; Gotelli and Colwell, 2001; Magurran, 2004).

Drawbacks regarding sample size can signify an important concern in environmental microbial studies. There are sensitivity variations in sampling size along the different diversity metrics. For example, estimators based on species richness are highly sensitive to sampling effort, e.g. Chao, Jackknife index (Hughes et al., 2001; Magurran, 2004). Certainly, others indices are found to be more accurate when measuring diversity at a low level of sampling density such as those based on taxonomic differences. In this regard, when the sampling effort is not exhaustive enough the accumulation curve can be 'extrapolated'.

An extrapolation of this curve allows predictions on the increase in species richness as the sampling effort is intensified rather than an estimate of total richness valuation (Magurran, 2004). But, it is argued that this technique has limited access in microbial studies since these communities are often very abundant/diverse and so the accumulation curve has either no yet begun to reach the asymptote or does not fit the best extrapolation model for predicting its level off (Lozupone, 2007). These issues are quite well controlled by counting with a deep sequencing dataset to increase the sampling effort as much as possible nearby the asymptote such as in the case of this study.

A statistical expectation of the corresponding accumulation curve is estimated by 'interpolation' processing most commonly referred as 'rarefaction [v] ' (Colwell et al., 2012). Rarefaction generates the 'expected' number of species in a small collection of *n* individuals from a larger pool of N individuals (the entire collection, i.e. the curve depends upon every individual in the pool at the accumulation curve's right-hand end). In opposition to accumulation curves, rarefaction curves move right to left, as the full dataset is increasingly rarefied (Gotelli and Colwell, 2001).

The purpose of generating rarefaction curves is to make direct comparisons among communities on the basis of a number of individuals in the smallest samples (Crist and Veech, 2006; Magurran, 2004). The method has been widely applied in microbial ecology, especially, to estimate the effectiveness of sampling effort to highly represent the diversity of the total

[v] *In spite of the controversy behind the terminology 'rarefaction' due to its original correspondence with another technique, I use this term instead of 'rarefying' since is the one referred in QIIME pipelines, the platform I applied when processing diversity analyses as described in methods (Chapter 3).*

microbial community, which is particularly critical when working on the basis of millions of DNA sequences distributed unevenly along all the species/OTUs. For example, it is unreliable analyze diversity in which some of the OTUs end up with millions of DNA sequences whereas others count with only one (singleton) or two (doubleton) sequences. For this reason, diversity is estimated using a rarefied dataset and all the measures are made using the same number of sequences, i.e. at the same depth of sampling.

McMurdie and Holmes, (2014) stated that rarefaction is one of the common procedures for addressing differences in sequencing effort across samples (different library sizes) – another classic one applies the proportional abundance of each species in a library. These authors are formal detractors of 'rarefaction' by arguing that it throw away data of the individuals from the larger libraries which is a waste of valuable information. Regardless, they highlighted that rarefaction are adequate when comparing 'obviously different' microbiomes such as in this investigation.

'Interpolation' (rarefaction) and 'extrapolation' has been more clearly explained in Colwell et al., (2012) who pointed out that an *interpolation* estimates the 'expected' number of species in a random sample of a smaller number of individuals or a smaller area sampled, meanwhile, *extrapolation* estimates the number of species that 'might be expected' in a larger number of individuals or a larger area samples.

Summarizing, either on the basis of raw-observed, normalized, extrapolated or rarefied data the final shape of the accumulation curve estimate diversity as well as the sampling effort effectiveness to represent such diversity. As indicated by Hughes et al., 2001, both richness and relative abundances differences in the sampled communities underlie the differences in the shape of the curves. Thereby, these curves: (i) scope the total diversity of the community that have been sampled, (ii) qualifies how copiously representative can be different depth of sampling to estimate the total diversity of the community, i.e. the effectiveness of the sampling effort (iii) the curve can be extrapolated to estimates the total species richness when the sampling effort is not sufficiently exhaustive and, (iv) a surrogate of this type of curve (rarefaction curves) represent the way how to compare among communities unevenly sampled, i.e. with totally different number of sequences (Crist and Veech, 2006; Hughes et al., 2001; Magurran, 2004).

Diversity measures according to spatial scales: alpha, beta, and gamma diversity

So far, I have reviewed what embodies 'biodiversity measures' however, what concern the use of them for local or comparative analysis has not been discussed yet. This is because, transversally, all the criteria described above somehow influence diversity measurements in both cases. Nevertheless, one of the most popular concerns of biological diversity studies is the proper partitioning of the communities when diversity is evaluated (e.g. into α-diversity and β-diversity) across space and/or time and any kind of comparative analysis (Magurran, 2004; Whittaker, 1972). Consequently, different metrics for local and comparative diversity analyses

have been developed, although framed on the same basis described above (e.g. weighing and unweighted by abundance/genetic relatedness) and influenced by same factors such as 'the sampling effort'.

Whittaker in 1972 proposed that estimation of biodiversity over a *spatial* scale require being hierarchized by partitioning the community into *alpha* (*α-diversity*), *beta* (*β-diversity*) and *gamma diversity* (*γ-diversity*) – even others (ε-diversity) (Table 1-2). On the basis of Whittaker's framework but as described by Magurran in 2004, *α-diversity* is the property of a defined spatial unit, while *β-diversity or 'turnover'* is a measure of the extent to which the diversity of two or more spatial units differ in terms of their species composition. Lozupone in 2007 '*whereas measurements of α-diversity can be used to compare the amount of diversity in different environmental samples, measurements of β-diversity are used to compare the type of diversity in different environments or along gradients*'. In other terms, Nemergut et al., in 2013 simply stated that α-*diversity* and β-*diversity* are respectively also referred as *inventory* and *differentiation* diversity.

More recently, Lynch and Neufeld in 2015 have described that α-diversity correspond with the richness of a specific community, i.e. how many species exist in a simple sample, site or habitat. On the other hand, β-diversity represent the differentiation of communities along different environments, i.e. how communities change across a range of samples, sites or habitats.

Table 1−2. Categories of inventory and differentiation diversity in relation to the scale of investigation (after Whittaker, 1972).

Scale	Inventory diversity	Differentiation diversity
Within sample	Point diversity	
Between samples, within habitat		Pattern diversity
Within habitat	α-diversity	
Between habitat, within landscape		β-diversity
Within landscape	γ-diversity	
Between landscape		δ-diversity
Within biogeographic province	ε-diversity	

In our particular case, α-diversity represent the diversity characterisation within each individual community of the forty-nine sampling sites/ecosystems composing $NS_{\text{-transect}}$ and $WE_{\text{-transect.}}$ And β-diversity community represents the change and differentiation distance among microbial communities diversity when comparing our sampling sites/ecosystems along and between the two transects. Finally, the total diversity of our entire study area represents the gamma diversity which results will lastly use to extend some prediction about microbial distribution across the entire NSW region.

In principle, any level of inventory and differential diversity (Table 1-2), such as *α-diversity* and *β-diversity,* count with estimators (e.g. rarefaction plots, indices) to characterize diversity from different perspectives, e.g. richness and evenness weighted/unweighted by genetic variations.

In this research, the soil microbial structural diversity has been assessed by characterizing richness and evenness of the microbial communities using a set of different estimators in a way to provide a more comprehensive and close interpretation of the real scenario. Some of the specific α-diversity and β-diversity (richness/evenness) metrics applied in our study are briefly reviewed in Table 1-3 and in Table 1-4, respectively.

Table 1-3. α-diversity estimators for richness and evenness characterisation based on both species-based and divergence-based methods (modified from Lozupone and Knight, 2008).

Test	Measurement	Application	Reference
	α-diversity richness (qualitative)		
Chao1	The total amount of observed species.	Useful for dataset skewed toward the low abundance groups.	Chao, 1984
Phylogenetic Diversity (PD)	The amount or proportion of branch length in a phylogenetic tree that leads to organisms from a community.	Determines which communities are the most phylogenetically diverse. (Does the phylogenetic richness of a community decrease with pollution or disease?)	Faith, 1992
	α-diversity evenness (quantitative)		
Shannon	The proportion of species relative to the total number of species. Emphasize richness component but account abundance.	It assumes that individuals are randomly sampled from an infinitely large community for which all the species are represented in that sample.	Shannon, 1948
Simpson	In essence, it captures the variance of the species abundance distribution. It is heavily weighted towards the most abundant species in the sample for which is less sensitive to species richness.	It gave the probability of any two individuals drawn at random from an infinitely large community belonging to the same species. It will rank species.	Simpson, 1949
Theta	The average divergence between 2 randomly chosen individuals in a community.	Determines how phylogenetically distinct individuals in a community are. (Does the phylogenetic evenness of a community decrease with pollution or disease?)	Martin, 2002

Such measurements were performed using QIIME and PIPIT pipelines. Each of those procedures and the specific configuration used in both *α-diversity* and *β-diversity* analyses is described in the next heading.

Table 1-4. β-diversity estimators for richness and evenness characterisation based on both species-based and divergence-based methods (modified from Lozupone and Knight, 2008).

Test	Measurement	Unique traits	Reference
		β-diversity richness (qualitative)	
Jaccard	Measure the number of species that are shared between two samples and the number that are unique to one sample or the other.	Qualitative measures that estimate similarity among communities. It is defined as the size of the shared species divided by the size of the total species from all the compared samples.	Jaccard, 1901
Unifrac	Significance: More unique evolution has occurred within the communities than expected by chance Clustering: Similar communities have similar phylogenetic lineages.	Qualitative measures that exclusively uses a phylogenetic tree and accounts for the history of shared ancestry between communities.	Lozupone and Knight, 2005
		β-diversity evenness (quantitative)	
Bray-Curtis	Measure number of species that are shared between two samples and the number that are unique to one sample or the other.	Quantitative measure which quantified the compositional dissimilarity among samples based on counts at each sample.	Bray and Curtis, 1957
Weighted unifrac	Significance: The individuals within the communities are more phylogenetically similar to each other than to those in another community. Clustering: Similar communities contain more phylogenetically similar individuals	Quantitative version of UniFrac. Similar to clustering with FST or DPCoA but exclusively uses a phylogenetic tree.	Lozupone et al., 2007

REFERENCES

Akondi, K.B., Lakshmi, V.V., 2013. Emerging Trends in Genomic Approaches for Microbial Bioprospecting. Omi. A J. Integr. Biol. 17, 61–70. doi:10.1089/omi.2012.0082

Allison, S.D., Martiny, J.B.H., 2008. Colloquium paper: resistance, resilience, and redundancy in microbial communities. Proc. Natl. Acad. Sci. U. S. A. 105 Suppl, 11512–11519. doi:10.1073/pnas.0801925105

Aschonitis, V.G., Castaldelli, G., Fano, E.A., 2016. Relations between environmental gradients and diversity indices of benthic invertebrates in lotic systems of northern Italy. Web Ecol. 16, 13–15. doi:10.5194/we-16-13-2016

Ayrapetyan, M., Oliver, J.D., 2016. The viable non-culturable state and its relevance in food safety. Curr. Opin. Food Sci. doi:10.1016/j.cofs.2016.04.010

Baas-Becking, L.G.M., 1934. Geobiologie of inleiding tot de milieukunde. Hague, Neth. van Stock. Zoon. 263.

Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J., Rídl, J., Vlček, Č., Voříšková, J., 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J. 6, 248–258. doi:10.1038/ismej.2011.95

Banwart, S., Noellemeyer, E., Milne, E., 2015. Soil carbon: Science, management, and policy for multiple benefits, Scope. CABI, Wallingford, UK, UK.

Blagodatskaya, E., Kuzyakov, Y., 2013. Active microorganisms in soil: Critical review of

estimation criteria and approaches. Soil Biol. Biochem. 67, 192–211. doi:10.1016/j.soilbio.2013.08.024

Bray, J.R., Curtis, J.T., 1957. An Ordination of the upland forest community of southern Wisconsin.pdf. Ecol. Monogr. doi:10.2307/1942268

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: history and future directions. Geoderma 264, 256–274.

Buckland, S.T., Magurran, a E., Green, R.E., Fewster, R.M., 2005. Monitoring change in biodiversity through composite indices. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360, 243–254. doi:10.1098/rstb.2004.1589

Burrough, P.A., Bouma, J., Yates, S.R., 1994. The state of the art in pedometrics. Geoderma 62, 311–326. doi:10.1016/0016-7061(94)90043-4

Cadotte, M.W., Jonathan Davies, T., Regetz, J., Kembel, S.W., Cleland, E., Oakley, T.H., 2010. Phylogenetic diversity metrics for ecological communities: Integrating species richness, abundance and evolutionary history. Ecol. Lett. 13, 96–105. doi:10.1111/j.1461-0248.2009.01405.x

Cao, P., Zhang, L.M., Shen, J.P., Zheng, Y.M., Di, H.J., He, J.Z., 2012. Distribution and diversity of archaeal communities in selected Chinese soils. FEMS Microbiol. Ecol. 80, 146–158. doi:10.1111/j.1574-6941.2011.01280.x

Chao, A., 1984. Nonparametric Estimation of the Number of Classes in a Population. Scand. J. Stat. 11, 265–270.

Chong, C.W., Pearce, D. a., Convey, P., Yew, W.C., Tan, I.K.P., 2012. Patterns in the distribution

of soil bacterial 16S rRNA gene sequences from different regions of Antarctica. Geoderma 181–182, 45–55. doi:10.1016/j.geoderma.2012.02.017

Colwell, R.K., 2013. EstimateS 9.1.0 User's Guide [WWW Document]. Estim. 9.1.0 User's Guid. URL>http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/EstimateSUsersGuide.htm (accessed 9.1.16).

Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L., Longino, J.T., 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. J. Plant Ecol. 5, 3–21. doi:10.1093/jpe/rtr044

Constancias, F., Terrat, S., Saby, N.P.A., Horrigue, W., Villerd, J., Guillemin, J.P., Biju-Duval, L., Nowak, V., Dequiedt, S., Ranjard, L., Chemidlin Pr??vost-Bour??, N., 2015. Mapping and determinism of soil microbial community distribution across an agricultural landscape. Microbiologyopen 505–517. doi:10.1002/mbo3.255

Crist, T.O., Veech, J.A., 2006. Additive partitioning of rarefaction curves and species-area relationships: Unifying ??-, ??- and ??-diversity with sample size and habitat area. Ecol. Lett. 9, 923–932. doi:10.1111/j.1461-0248.2006.00941.x

Crits-Christoph, A., Robinson, C.K., Barnum, T., Fricke, W.F., Davila, A.F., Jedynak, B., McKay, C.P., Diruggiero, J., 2013. Colonization patterns of soil microbial communities in the Atacama Desert. Microbiome 1, 28. doi:10.1186/2049-2618-1-28

Daniel, R., 2005. The metagenomics of soil. Nat. Rev. Microbiol. 3, 470–478. doi:10.1038/nrmicro1160

de Gannes, V., Eudoxie, G., Bekele, I., Hickey, W.J., 2015. Relations of microbiome characteristics

to edaphic properties of tropical soils from Trinidad. Front. Microbiol. 6, 1–13. doi:10.3389/fmicb.2015.01045

De Groot, R.S., Wilson, M. a., Boumans, R.M.J., 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. 41, 393–408. doi:10.1016/S0921-8009(02)00089-7

de Vries, F.T., Manning, P., Tallowin, J.R.B., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Hobbs, P.J., Quirk, H., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bardgett, R.D., 2012. Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. Ecol. Lett. 15, 1230–1239. doi:10.1111/j.1461-0248.2012.01844.x

Dequiedt, S., Thioulouse, J., Jolivet, C., Saby, N.P. a, Lelievre, M., Maron, P.A., Martin, M.P., Prévost-Bouré, N.C., Toutain, B., Arrouays, D., Lemanceau, P., Ranjard, L., 2009. Biogeographical patterns of soil bacterial communities. Environ. Microbiol. Rep. 1, 251–255. doi:10.1111/j.1758-2229.2009.00040.x

Dimitriu, P.A., Grayston, S.J., 2010. Relationship between soil properties and patterns of bacterial β-diversity across reclaimed and natural boreal forest soils. Microb. Ecol. 59, 563–573. doi:10.1007/s00248-009-9590-0

Dokic, L., Savic, M., Narancic, T., Vasiljevic, B., 2010. Metagenomic Analysis of Soil Microbial Communities. Arch. Biol. Sci. Belgrade 62, 559–564. doi:10.2298/ABS1003559J

Dokuchaev, V. V, 1883. The Russian Chernozem Report to the Free Economic Society. Imp. Univ. St. Petersburg, St. Petersburg, Russ.

Elshahed, M.S., Youssef, N.H., Spain, A.M., Sheik, C., Najar, F.Z., Sukharnikov, L.O., Roe, B.A., Davis, J.P., Schloss, P.D., Bailey, V.L., Krumholz, L.R., 2008. Novelty and uniqueness

patterns of rare members of the soil biosphere. Appl. Environ. Microbiol. 74, 5422–5428. doi:10.1128/AEM.00410-08

Escarguel, G., Fara, E., Brayard, A., Legendre, S., 2011. Biodiversity is not (and never has been) a bed of roses! Comptes Rendus - Biol. 334, 351–359. doi:10.1016/j.crvi.2011.02.006

Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10. doi:10.1016/0006-3207(92)91201-3

Fajardo, M., McBratney, A., Whelan, B., 2016a. Fuzzy clustering of Vis-NIR spectra for the objective recognition of soil morphological horizons in soil profiles. Geoderma 263, 244–253. doi:10.1016/j.geoderma.2015.05.010

Fajardo, M., McBratney, A.B., Field, D.J., Minasny, B., 2016b. Soil slaking assessment using image recognition. Soil Tillage Res. 163, 119–129. doi:10.1016/j.still.2016.05.018

Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. Ecol. Econ. 68, 643–653. doi:10.1016/j.ecolecon.2008.09.014

Fuhrman, J. a, 2009. Microbial community structure and its functional implications. Nature 459, 193–199. doi:10.1038/nature08058

Gasc, C., Ribière, C., Parisot, N., Beugnot, R., Defois, C., Petit-Biderre, C., Boucher, D., Peyretaillade, E., Peyret, P., 2015. Capturing prokaryotic dark matter genomes. Res. Microbiol. 166. doi:10.1016/j.resmic.2015.06.001

Gest, H., 2004. The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. Notes Rec. R. Soc. 58, 187–201.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J., 2005. Re-evaluating prokaryotic

species. Nat.Rev.Microbiol. 3, 733–739. doi:10.1038/nrmicro1236

Gotelli, N.J., Colwell, R.K., 2001. Quantifyinf biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol. Lett. 4 Abstract, 379–391. doi:10.1046/j.1461-0248.2001.00230.x

Green, B.D., Keller, M., 2006. Capturing the uncultivated majority. Curr. Opin. Biotechnol. 17, 236–240. doi:10.1016/j.copbio.2006.05.004

Griffiths, B.S., Kuan, H.L., Ritz, K., Glover, L. a., McCaig, a. E., Fenwick, C., 2004. The relationship between microbial community structure and functional stability, tested experimentally in an upland pasture soil. Microb. Ecol. 47, 104–113. doi:10.1007/s00248-002-2043-7

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils. Environ. Microbiol. 13, 1642–1654. doi:10.1111/j.1462-2920.2011.02480.x

Gugerli, F., Brandl, R., Castagneyrol, B., Martin, F., Peter, M., Pritsch, K., Marinus, J.M., 2013. Community genetics in the time of next- generation molecular technologies 41, 3198–3207.

Handelsman, J., 2004. Metagenomics : Application of Genomics to Uncultured Microorganisms. Microbiol. Mol. Biol. Rev. 68, 669–685. doi:10.1128/MBR.68.4.669

Hartemink, A.E., Minasny, B., 2014. Towards digital soil morphometrics. Geoderma 230–231, 305–317. doi:10.1016/j.geoderma.2014.03.008

Hazen, T.C., Rocha, A.M., Techtmann, S.M., 2013. Advances in monitoring environmental microbes. Curr. Opin. Biotechnol. 24, 526–533. doi:10.1016/j.copbio.2012.10.020

Henry, C.S., Overbeek, R., Xia, F., Best, A. a., Glass, E., Gilbert, J., Larsen, P., Edwards, R., Disz,

T., Meyer, F., Vonstein, V., Dejongh, M., Bartels, D., Desai, N., D'Souza, M., Devoid, S., Keegan, K.P., Olson, R., Wilke, A., Wilkening, J., Stevens, R.L., 2011. Connecting genotype to phenotype in the era of high-throughput sequencing. Biochim. Biophys. Acta - Gen. Subj. 1810, 967–977. doi:10.1016/j.bbagen.2011.03.010

Hill, T.C.J., Walsh, K.A., Harris, J.A., Moffet, B.F., 2003. Using ecological diversity measures with bacterial communities. FEMS Microbiol. Ecol. 43, 1–11. doi:10.1016/S0168-6496(02)00449-X

Howieson, J.G., Dilworth (Eds.), M.J., 2016. Working with rhizobia 314.

Hu, J., Lin, X., Wang, J., Dai, J., Chen, R., Zhang, J., Wong, M.H., 2011. Microbial functional diversity, metabolic quotient, and invertase activity of a sandy loam soil as affected by long-term application of organic amendment and mineral fertilizer. J. Soils Sediments 11, 271–280. doi:10.1007/s11368-010-0308-1

Huang, J., Sheng, X.F., Xi, J., He, L.Y., Huang, Z., Wang, Q., Zhang, Z.D., 2014. Depth-related changes in community structure of culturable mineral weathering bacteria and in weathering patterns caused by them along two contrasting soil profiles. Appl. Environ. Microbiol. 80, 29–42. doi:10.1128/AEM.02335-13

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., Bohannan, B.J.M., 2001. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. Appl. Environ. Microbiol. 67, 4399–4406. doi:10.1128/AEM.67.10.4399-4406.2001

Ibáñez, J.J., Feoli, E., 2013. Global Relationships of Pedodiversity and Biodiversity. Vadose Zo. J. 0, 0. doi:10.2136/vzj2012.0186

Izquierdo, J. a., Nüsslein, K., 2006. Distribution of extensive nifH gene diversity across physical

soil microenvironments. Microb. Ecol. 51, 441–452. doi:10.1007/s00248-006-9044-x

Jaccard, P., 1901. Etude de la distribution florale dans une portion des Alpes et du Jura. Bull. la Soc. Vaudoise des Sci. Nat. 37, 547–579. doi:10.5169/seals-266450

Jenny, H., 1941. Factors of soil formation; a sytem of quantitative pedology (techreport).

Joseph, S.J., Hugenholtz, P., Sangwan, P., Osborne, C.A., Janssen, P.H., 2003. Laboratory Cultivation of Widespread and Previously Uncultured Soil Bacteria. Appl. Environ. Microbiol. 69, 7210–7215. doi:10.1128/AEM.69.12.7210-7215.2003

Kasel, S., Bennett, L.T., Tibbits, J., 2008. Land use influences soil fungal community composition across central Victoria, south-eastern Australia. Soil Biol. Biochem. 40, 1724–1732. doi:10.1016/j.soilbio.2008.02.011

Konstantinidis, K.T., Ramette, A., Tiedje, J.M., 2006. The bacterial species definition in the genomic era. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 361, 1929–40. doi:10.1098/rstb.2006.1920

Krause, S., Meima-Franke, M., Hefting, M.M., Bodelier, P.L.E., 2013. Spatial patterns of methanotrophic communities along a hydrological gradient in a riparian wetland. FEMS Microbiol. Ecol. 86, 59–70. doi:10.1111/1574-6941.12091

Kuang, J.-L.L., Huang, L.-N.N., Chen, L.-X.X., Hua, Z.-S.S., Li, S.-J.J., Hu, M., Li, J.-T.T., Shu, W.-S.S., 2013. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. ISME J. 7, 1038–50. doi:10.1038/ismej.2012.139

Larsen, P., Hamada, Y., Gilbert, J., 2012. Modeling microbial communities: Current, developing, and future technologies for predicting microbial community interaction. J. Biotechnol. 160, 17–24. doi:10.1016/j.jbiotec.2012.03.009

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75, 5111–5120. doi:10.1128/AEM.00335-09

Lauber, C.L., Strickland, M.S., Bradford, M. a., Fierer, N., 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol. Biochem. 40, 2407–2415. doi:10.1016/j.soilbio.2008.05.021

Liu, J., Hua, Z.S., Chen, L.X., Kuang, J.L., Li, S.J., Shu, W.S., Huang, L.N., 2014. Correlating microbial diversity patterns with geochemistry in an extreme and heterogeneous environment of mine tailings. Appl. Environ. Microbiol. 80, 3677–3686. doi:10.1128/AEM.00294-14

Liu, Z., Liu, G., Fu, B., Zheng, X., 2008. Relationship between plant species diversity and soil microbial functional diversity along a longitudinal gradient in temperate grasslands of Hulunbeir, Inner Mongolia, China. Ecol. Res. 23, 511–518. doi:10.1007/s11284-007-0405-9

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., Vargas, C., Lima-Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels-Lewis, S., Karsenti, E., Bork, P., Acinas, S.G., 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environ. Microbiol. 16, 2659–2671. doi:10.1111/1462-2920.12250

Loman, N.J., Pallen, M.J., 2015. Twenty years of bacterial genome sequencing. Nat. Rev. Microbiol. 13, 1–9. doi:10.1038/nrmicro3565

Lombard, N., Prestat, E., van Elsas, J.D., Simonet, P., 2011. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. FEMS Microbiol. Ecol. 78, 31–49. doi:10.1111/j.1574-6941.2011.01140.x

Lozupone, C.A., 2007. Global patterns in bacterial diversity. ProQuest Diss. Theses Glob. University of Colorado.

Lozupone, C.A., Hamady, M., Kelley, S.T., Knight, R., 2007. Quantitative and qualitative ?? diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol. 73, 1576–1585. doi:10.1128/AEM.01996-06

Lozupone, C.A., Knight, R., 2008. Species Divergence and the Measurement of Microbial Diversity. Micorbiology Rev. 32, 557–578. doi:10.1111/j.1574-6976.2008.00111.x.Species

Lozupone, C., Knight, R., 2005. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. Appl. Environ. Microbiol. 71, 8228–8235. doi:10.1128/AEM.71.12.8228

Lynch, M.D.J., Neufeld, J.D., 2015. Ecology and exploration of the rare biosphere. Nat. Rev. Microbiol. 13, 217–229. doi:10.1038/nrmicro3400

M.J. Swift M. van Noordwijk, A.-M.N.I., Swift, M.J., Izac, a. M.N., Van Noordwijk, M., M.J. Swift M. van Noordwijk, A.-M.N.I., Swift, M.J., Izac, a. M.N., Van Noordwijk, M., 2004. Biodiversity and ecosystem services in agricultural landscapes - Are we asking the right questions? Agric. Ecosyst. Environ. 104, 113–134. doi:10.1016/j.agee.2004.01.013

Maestre, F.T., Delgado-Baquerizo, M., Jeffries, T.C., Eldridge, D.J., Ochoa, V., Gozalo, B., Quero, J.L., García-Gómez, M., Gallardo, A., Ulrich, W., Bowker, M.A., Arredondo, T., Barraza-Zepeda, C., Bran, D., Florentino, A., Gaitán, J., Gutiérrez, J.R., Huber-Sannwald, E., Jankju,

M., Mau, R.L., Miriti, M., Naseri, K., Ospina, A., Stavi, I., Wang, D., Woods, N.N., Yuan, X., Zaady, E., Singh, B.K., 2015. Increasing aridity reduces soil microbial diversity and abundance in global drylands. Proc. Natl. Acad. Sci. U. S. A. 112, 15684–15689. doi:10.1073/pnas.1516684112

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Publishing Ltd, Oxford, UK.

Martin, A.P., 2002. Phylogenetic Approaches for Describing and Comparing the Diversity of Microbial Phylogenetic Approaches for Describing and Comparing the Diversity of Microbial Communities. DNA Seq. 68, 3673–3682. doi:10.1128/AEM.68.8.3673

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J. a, Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreås, L., Reysenbach, A.-L., Smith, V.H., Staley, J.T., 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. 4, 102–112. doi:10.1038/nrmicro1341

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. doi:10.1016/j.geoderma.2013.08.013

McBratney, A., Minasny, B., Pino, V., Padarian, J., Fajardo, M., 2015. BIOPEDOMETRICS: THE PARALLELS CHALLENGES AND SYNERGIES OF SOIL BIODIVERSITY AND PEDODIVERSITY. Pedometrics 2015 173.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping, Geoderma. doi:10.1016/S0016-7061(03)00223-4

McBratney, a B., Minasny, B., Wheeler, I., Malone, B.P., 2012. Frameworks for digital soil assessment. Digit. Soil Assessments Beyond 9–14.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M.,

Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P., 2007. Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. Ecol. Lett. 10, 995–1015. doi:10.1111/j.1461-0248.2007.01094.x

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput. Biol. 10. doi:10.1371/journal.pcbi.1003531

Mendoza, M.L.Z., Sicheritz-Pont??n, T., Thomas Gilbert, M.P., 2014. Environmental genes and genomes: Understanding the differences and challenges in the approaches and software for their analyses. Brief. Bioinform. 16, 745–758. doi:10.1093/bib/bbv001

Millennium Ecosystem Assessment, Walter V. Reid Angela Cropper, Doris Capistrano, Stephen R. Carpenter, Kanchan Chopra, H.A.M., Partha Dasgupta Anantha Kumar Duraiappah, Rashid Hassan, Roger Kasperson, Rik Leemans, T.D., Robert M. May Prabhu Pingali, Cristián Samper, Robert Scholes, Robert T. Watson, T. (A. J.. M., A.H. Zakri Neville J. Ash, Elena Bennett, Pushpam Kumar, Marcus J. Lee, Ciara Raudsepp-Hearne, Z.S., Henk Simons and Monika B. Zurek, J.T., 2005. <Ecosystems and human well-being : synthesis / Millennium Ecosystem Assessment.>. Millenn. Ecosyst. Assess. Ser. 155.

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., M??ller, C., Obermaier, E., Prati, D., Socher, S.A., Sonnemann, I., W??schke, N., Wubet, T., Wurst, S., Rillig, M.C., 2014. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. Ecol. Evol. 4, 3514–3524. doi:10.1002/ece3.1155

Nahlik, A.M., Kentula, M.E., Fennessy, M.S., Landers, D.H., 2012. Where is the consensus? A

proposed foundation for moving ecosystem service concepts into practice. Ecol. Econ. 77, 27–35. doi:10.1016/j.ecolecon.2012.01.001

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L., Knight, R., 2011. Global patterns in the biogeography of bacterial taxa. Environ. Microbiol. 13, 135–144. doi:10.1111/j.1462-2920.2010.02315.x

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F., Knelman, J.E., Darcy, J.L., Lynch, R.C., Wickey, P., Ferrenberg, S., 2013. Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. 77, 342–56. doi:10.1128/MMBR.00051-12

Nesme, J., Achouak, W., Agathos, S.N., Bailey, M., Baldrian, P., Heulin, T., Jansson, J.K., Jurkevitch, E., Kruus, K., Kowalchuk, G.A., Lagares, A., Lappin-scott, H., Lemanceau, P., Le, D., Mandic-mulec, I., Murrell, J.C., Myrold, D.D., Nalin, R., Nannipieri, P., Neufeld, J.D., Parnell, J.J., Pylro, V., Ramos, J.L., Roesch, F.W., Schloter, M., Schleper, C., Sczyrba, A., Sessitsch, A., Tebbe, C., Topp, E., Tsiamis, G., Elsas, J.D. Van, Keulen, G. Van, Widmer, F., Wagner, M., Zhang, T., Zhang, X., Zhao, L., Zhu, Y., Vogel, T.M., Simonet, P., Genomics, E.M., Lyon, E.C. De, Environment, N., Road, M., Science, F., Directorate, C.S., Biology, E., Plata, L., Plata, L., Science, E., Science, B., Ecological, N., Network, O., Gabriel, S., 2016. Back to the future of soil metagenomics. Front. Microbiol. doi:10.3389/fmicb.2016.00073

Nottingham, A., Fierer, N., Turner, B., Whitaker, J., Ostle, N., McNamara, N., Bardgett, R., Leff, J., Salinas, N., Ccahuana, A., others, 2016. Temperature drives plant and soil microbial diversity patterns across an elevation gradient from the Andes to the Amazon. bioRxiv

79996.

Nunan, N., Wu, K., Young, I.M., Crawford, J.W., Ritz, K., 2002. In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. Microb. Ecol. 44, 296–305. doi:10.1007/s00248-002-2021-0

Oline, D.K., Schmidt, S.K., Grant, M.C., 2006. Biogeography and landscape-scale diversity of the dominant Crenarchaeota of soil. Microb. Ecol. 52, 480–490. doi:10.1007/s00248-006-9101-5

Orgiazzi, A., Bardgett, R.D., Barrios, E., Behan-Pelletier, V., B., M.J.I., Chotte, J-L., De Deyn, G.B., Eggleton, P., Fierer, N., Fraser, T., Hedlund, K., Jeffery, S., Johnson, N.C., Jones, A., Kandeler, E., Kaneko, N., Lavelle, P., Lemanceau, P., M, D.., Orgiazzi, A., Bardgett, R.D., Barrios, E., Behan-Pelletier, V., Briones, M.J.I., Chotte, J.-L., De Deyn, G.B., Eggleton, P., Fierer, N., Fraser, T., Others, 2016. Global Soil Biodiversity Atlas, European Commission, Publications. Publications Office of the European Union, Luxembourg. doi:10.2788/2613

Pedrós-Alió, C., 2012. The Rare Bacterial Biosphere. Ann. Rev. Mar. Sci. 4, 449–466. doi:10.1146/annurev-marine-120710-100948

Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N.P. a, Lelievre, M., Maron, P. a, Morin, F.E.R., Bispo, a, Jolivet, C., Arrouays, D., Lemanceau, P., 2013. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. Nat. Commun. 4, 1434. doi:10.1038/ncomms2431

Ranjard, L., Lejon, D.P.H., Mougel, C., Schehrer, L., Merdinoglu, D., Chaussod, R., 2003. Sampling strategy in molecular microbial ecology: Influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environ. Microbiol. 5, 1111–

1120. doi:10.1046/j.1462-2920.2003.00521.x

Rastogi, G., Sani, R.K., 2011. Molecular techniques to assess microbial community structure, function, and dynamics in the environment. Microbes Microb. Technol. Agric. Environ. Appl. 29–57. doi:10.1007/978-1-4419-7931-5_2

Reid, A., Buckley, M., Mcfall-, M., 2011. The Rare Biosphere. Am. Acad. Microbiol.

Reuter, J.A., Spacek, D. V., Snyder, M.P., 2015. High-Throughput Sequencing Technologies. Mol. Cell 58, 586–597. doi:10.1016/j.molcel.2015.05.004

Richter, I., Herbold, C.W., Lee, C.K., McDonald, I.R., Barrett, J.E., Cary, S.C., 2014. Influence of soil properties on archaeal diversity and distribution in the McMurdo Dry Valleys, Antarctica. FEMS Microbiol. Ecol. 89, 347–359. doi:10.1111/1574-6941.12322

Rime, T., Hartmann, M., Brunner, I., Widmer, F., Zeyer, J., Frey, B., 2015. Vertical distribution of the soil microbiota along a successional gradient in a glacier forefield. Mol. Ecol. n/a-n/a. doi:10.1111/mec.13051

Robeson, M.S., King, A.J., Freeman, K.R., Birky, C.W., Martin, A.P., Schmidt, S.K., 2011. Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. Proc. Natl. Acad. Sci. U. S. A. 108, 4406–4410. doi:10.1073/pnas.1012678108

Robinson, D. a., Hockley, N., Cooper, D.M., Emmett, B. a., Keith, a. M., Lebron, I., Reynolds, B., Tipping, E., Tye, a. M., Watts, C.W., Whalley, W.R., Black, H.I.J., Warren, G.P., Robinson, J.S., 2013. Natural capital and ecosystem services, developing an appropriate soils framework as a basis for valuation. Soil Biol. Biochem. 57, 1023–1033. doi:10.1016/j.soilbio.2012.09.008

Rousk, J., Baath, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R., Fierer,

N., 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. Isme J. 4, 1340–1351. doi:10.1038/ismej.2010.58

Schloss, P.D., Handelsman, J., 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol. 6, 229. doi:10.1186/gb-2005-6-8-229

Schmidt, P.A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina metabarcoding of a soil fungal community. Soil Biol. Biochem. 65, 128–132. doi:10.1016/j.soilbio.2013.05.014

Schneider, T., Riedel, K., 2010. Environmental proteomics: Analysis of structure and function of microbial communities. Proteomics 10, 785–798. doi:10.1002/pmic.200900450

Schulz, S., Brankatschk, R., Dümig, a., Kögel-Knabner, I., Schloter, M., Zeyer, J., 2013. The role of microorganisms and plants at different stages of ecosystem development for soil formation. Biogeosciences Discuss. 10, 3983–3996. doi:10.5194/bgd-10-1867-2013

Shahbazi, F., Aliasgharzad, N., Ebrahimzad, S. a., Najafi, N., 2013. Geostatistical analysis for predicting soil biological maps under different scenarios of land use. Eur. J. Soil Biol. 55, 20–27. doi:10.1016/j.ejsobi.2012.10.009

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. doi:10.1145/584091.584093

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794–1805. doi:10.1111/j.1365-294X.2012.05538.x

Simon, C., Daniel, R., 2011. Metagenomic analyses: Past and future trends. Appl. Environ. Microbiol. 77, 1153–1161. doi:10.1128/AEM.02345-10

Simpson, E.H., 1949. Measurement of Diversity. Nature 163, 688–688. doi:10.1038/163688a0

Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: Evaluating the illumina platform. PLoS One 10, 1–18. doi:10.1371/journal.pone.0116955

Singh, B.K., Quince, C., Macdonald, C. a., Khachane, A., Thomas, N., Al-Soud, W.A., Sørensen, S.J., He, Z., White, D., Sinclair, A., Crooks, B., Zhou, J., Campbell, C.D., 2014. Loss of microbial diversity in soils is coincident with reductions in some specialized functions. Environ. Microbiol. 16, 2408–2420. doi:10.1111/1462-2920.12353

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31. doi:10.1016/j.still.2004.03.008

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc. Natl. Acad. Sci. 103, 12115–12120. doi:10.1073/pnas.0605127103

Steven, B., Gallegos-Graves, L.V., Belnap, J., Kuske, C.R., 2013. Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. FEMS Microbiol. Ecol. 86, 101–113. doi:10.1111/1574-6941.12143

Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. Mol. Ecol. 21, 1789–1793. doi:10.1111/j.1365-294X.2012.05542.x

Tanaka, T., Kawasaki, K., Daimon, S., Kitagawa, W., Yamamoto, K., Tamaki, H., Tanaka, M., Nakatsu, C.H., Kamagata, Y., 2014. A hidden pitfall in the preparation of agar media undermines microorganism cultivability. Appl. Environ. Microbiol. 80, 7659–7666.

doi:10.1128/AEM.02741-14

Tkacz, A., Cheema, J., Chandra, G., Grant, A., Poole, P.S., 2015. Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. The ISME Journal 9, 2349–2359. doi:10.1038/ismej.2015.41

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, S., Wardle, D.A., Lindahl, B.D., 2014. Global diversity and geography of soil fungi. Science (80-.). 346, 1052–1053. doi:10.1126/science.aaa1185

Torsvik, V., Daae, F.L., Sandaa, R.A., Øvreås, L., 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. J. Biotechnol. 64, 53–62. doi:10.1016/S0168-1656(98)00103-5

Torsvik, V.L., 1980. Isolation of bacterial DNA from soil. Soil Biol. Biochem. 12, 15–21. doi:10.1016/0038-0717(80)90097-8

Torsvik L., V.O., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5, 240–245. doi:http://dx.doi.org/10.1016/S1369-5274(02)00324-7

Trangmar, B.B., Yost, R.S., Uehara, G., 1986. Application of geostatistics to spatial studies of soil properties. Adv. Agron. 38, 45–94.

Tuomisto, H., 2010. A consistent terminology for quantifying species diversity? Yes, it does exist. Oecologia 164, 853–860. doi:10.1007/s00442-010-1812-0

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G. a., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev. 37, 936–954. doi:10.1111/1574-6976.12023

Wagg, C., Bender, S.F., Widmer, F., van der Heijden, M.G.A., 2014. Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proceedings of the

National Academy of Sciences of the United States of America 111. doi:10.1073/pnas.1320054111

Wakelin, S. a., Gupta, V.V.S.R., Forrester, S.T., 2010. Regional and local factors affecting diversity, abundance and activity of free-living, N2-fixing bacteria in Australian agricultural soils. Pedobiologia (Jena). 53, 391–399. doi:10.1016/j.pedobi.2010.08.001

Webster, R., Oliver, M.A., others, 1990. Statistical methods in soil and land resource survey. Oxford University Press (OUP).

Whittaker, A.R.H., 1972. Evolution and Measurement of Species Diversity Published by : International Association for Plant Taxonomy (IAPT) Stable URL : http://www.jstor.org/stable/1218190 REFERENCES Linked references are available on JSTOR for this article : You may need to log 21, 213–251.

Xi, N., Bloor, J.M.G., 2016. Interactive effects of precipitation and nitrogen spatial pattern on carbon use and functional diversity in soil microbial communities. Appl. Soil Ecol. 100, 207–210. doi:10.1016/j.apsoil.2015.11.030

Xia, Z., Bai, E., Wang, Q., Gao, D., Zhou, J., Jiang, P., Wu, J., 2016. Biogeographic Distribution Patterns of Bacteria in Typical Chinese Forest Soils. Front. Microbiol. 7, 1106. doi:10.3389/fmicb.2016.01106

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., Yang, Y., Yao, T., Knight, R., Chu, H., 2012. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. Environ. Microbiol. 14, 2457–2466. doi:10.1111/j.1462-2920.2012.02799.x

Xu, Z., Hansen, M.A., Hansen, L.H., Jacquiod, S., Sørensen, S.J., 2014. Bioinformatic approaches

reveal metagenomic characterization of soil microbial community. PLoS One 9, e93445. doi:10.1371/journal.pone.0093445

Yasir, M., Azhar, E.I., Khan, I., Bibi, F., Baabdullah, R., Al-Zahrani, I. a, Al-Ghamdi, A.K., 2015. Composition of soil microbiome along elevation gradients in southwestern highlands of Saudi Arabia. BMC Microbiol. 15, 1–9. doi:10.1186/s12866-015-0398-4

Zarraonaindia, I., Smith, D.P., Gilbert, J. a., 2013. Beyond the genome: Community-level analysis of the microbial world. Biol. Philos. 28, 261–282. doi:10.1007/s10539-012-9357-8

Zhang, B., Liang, C., He, H., Zhang, X., 2013. Variations in Soil Microbial Communities and Residues Along an Altitude Gradient on the Northern Slope of Changbai Mountain, China. PLoS One 8. doi:10.1371/journal.pone.0066184

Zhang, Y., Cong, J., Lu, H., Li, G., Xue, Y., Deng, Y., Li, H., Zhou, J., Li, D., 2015. Soil bacterial diversity patterns and drivers along an elevational gradient on Shennongjia Mountain, China. Microb. Biotechnol. 8, 739–746. doi:10.1111/1751-7915.12288

Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., Qin, Y., Xue, K., Wu, L., He, Z., Voordeckers, J.W., Nostrand, J.D. Van, Buzzard, V., Michaletz, S.T., Enquist, B.J., Weiser, M.D., Kaspari, M., Waide, R., Yang, Y., Brown, J.H., 2016. Temperature mediates continental-scale diversity of microbes in forest soils. Nat. Commun. 7, 12083. doi:10.1038/ncomms12083

Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., Alvarez-cohen, L., 2015. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats 6, 1–17. doi:10.1128/mBio.02288-14.Copyright

Chapter 2

Designing a Sampling Scheme for Microbial Diversity Analysis in New South Wales

Outlines

The target area of this investigation extends along two transects, which together represent a longitudinal (north-south) and latitudinal (west-east) agroecological gradient across the state of New South Wales (NSW) in Australia. This chapter details the protocol applied to schematize the study area, the sampling sites and the collection of the soil samples for the evaluation of the soil microbial diversity along both NS-transect and WE-transect. The environmental conditions faced by these particular microbial communities within the boundaries of NSW are described at the beginning of this chapter.

Summary

Two transects were designed across the State of NSW in Australia, to represent a longitudinal (north-south) and latitudinal (east-west) agroecological gradient at large-scale. The north-south (NS) transect has been designed to encompass the 550 mm mean annual rainfall isohyet and extends approximately 900 km in length from Queensland to the Victorian border. Running perpendicular, the west-east (WE) transect extends approximately 930 km in length and follows mean annual rainfall gradients of >1500 mm to <300 mm; starting at Wanaaring in western NSW and ending in the area of Coffs Harbour in the coast (Figure 2-1).

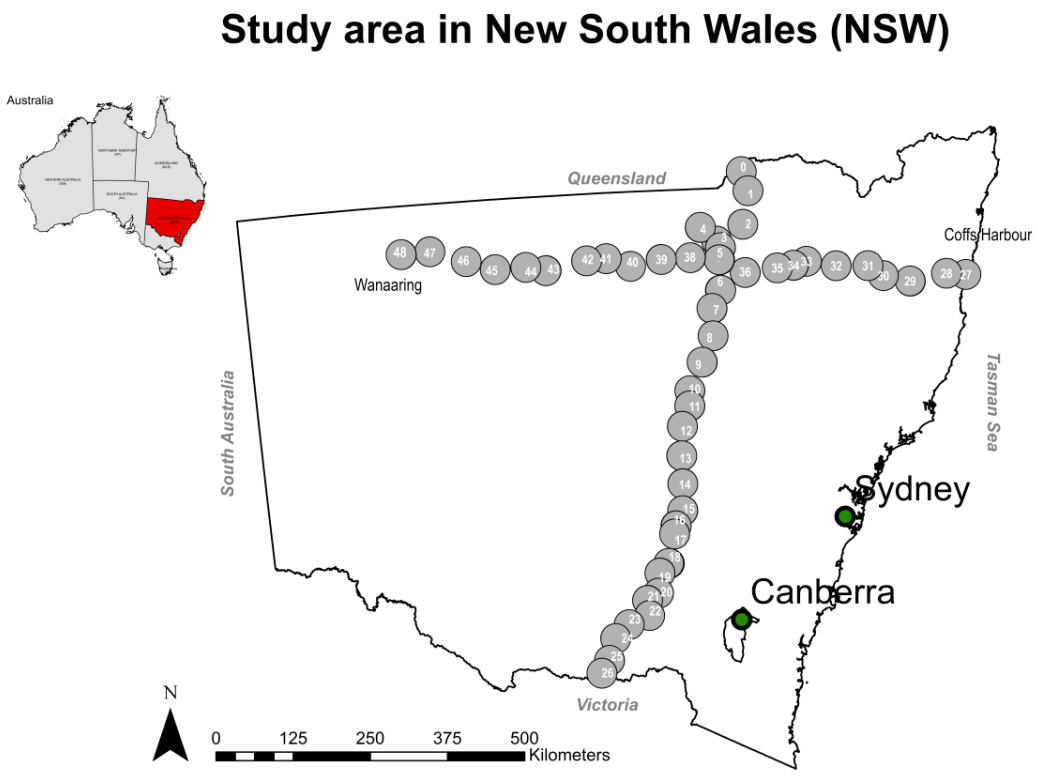


Figure 2−1. Study area represented by a longitudinal (NS-$_{transect}$) and latitudinal (WE-$_{transect}$) agroecological gradient across the state of NSW (Australia).

Based on the construction of a geographic information system, forty-nine sampling sites were designed and situated over these two transects at a separation distance of 50 km. Representative environmental areas of each sampling site (based on soil type, land use and others variables) were selected from two different land-use ecosystems. These land-use ecosystems were both natural (forest or grassland) and rainfed agriculture (crop or pasture). Soil samples intended to microbial and physicochemical analysis were collected from these individual ecosystems at each site at 0-5 cm and 5-10 cm depth.

More information regarding the environmental conditions of the study area, the design process of the sampling sites and the protocol used to collect the soil samples, are detailed in this chapter.

NEW SOUTH WALES IN THE CONTEXT OF THE AUSTRALIAN CONTINENT

New South Wales (NSW) is one of the six states composing Australia's federation and it possesses distinctive characteristics that make it an interesting area in which to explore soil microbial diversity at landscape scale. First, this state spans an immense total land area (800,642 km^2) which, in term of comparison, is slightly larger than France (643,801 km^2) (Figure 2-2) and six times larger than England (130,279 km^2). Secondly, NSW exhibits a great diversity of landscapes and habitats which, moreover, have a remarkable stability due to its geological nature (ABARES, 2012; EPA, 2012; Morton et al., 2014; NARCliM, 2014). Thirdly, a significant portion of one of Australia's most important agricultural areas take place in this area: *the Wheat-belt East Region* (Figure 2-3) (ABARES, 2012; EPA, 2012; Morton et al., 2014; NARCliM, 2014). In sum, all these natural conditions featuring NSW, and adding the fact that they have recently been well documented, make it an advantageous area in which to carry out this exploration and evaluation of microbial diversity distribution patterns on a wide scale in Australian soils.

NSW attributes - dimension, diversity and stability - as well as the environmental variables defining them (*e.g.* geomorphology, climate, soil, vegetation etc.), are intimately linked to the heritage of Australia's geological evolution since it was part of the Gondwana super-continent around 5.3 to 23 million years ago (Blewett, 2012; McKenzie et al., 2004; Morton et al., 2014). In order to deepen our understanding of these features which are specific to NSW, it is necessary to frame this particular study within a wider Australian context.

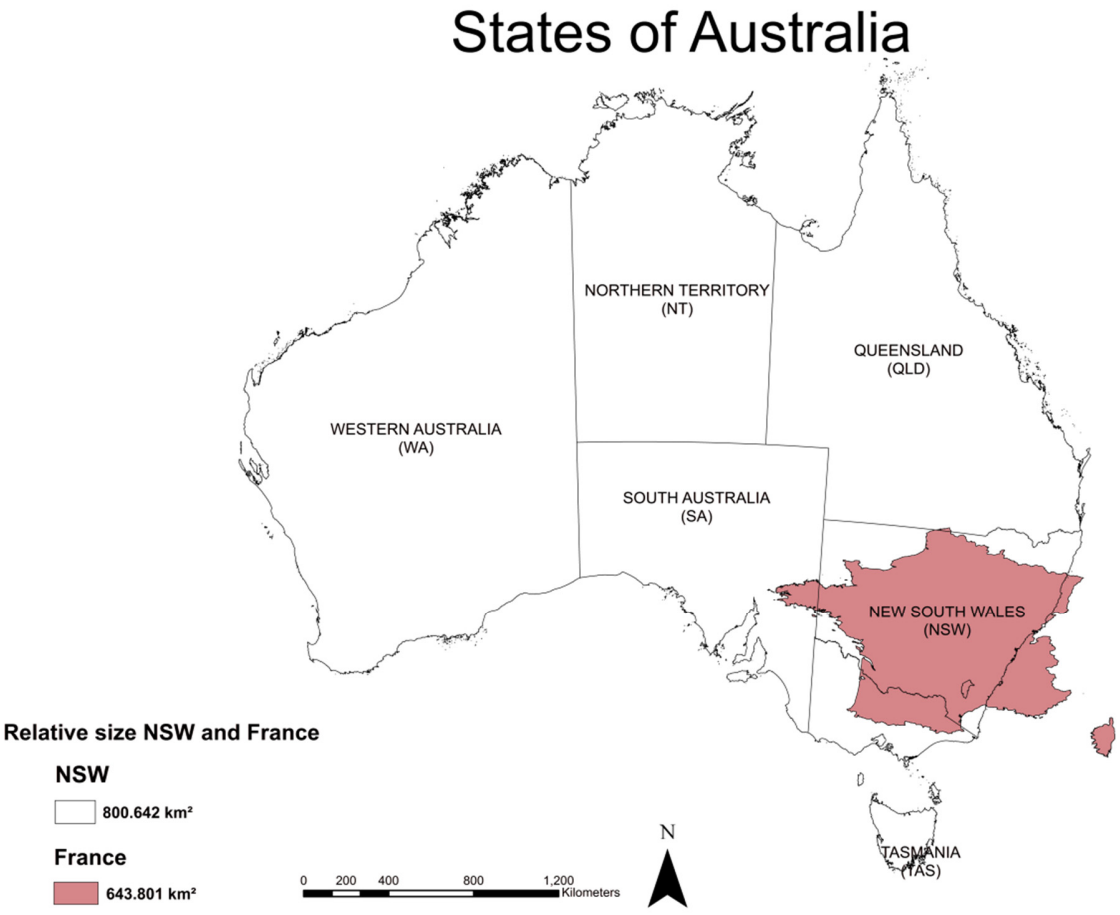


Figure 2-2. New South Wales (NSW) and its relative size to France.

Australia is one of the oldest, driest and most stable land surfaces on Earth. Amazingly when we actually compare this continent with the age of the Earth, we find that Australia is only 150 million years younger and has been stable for the past 200 million years (Blewett, 2012). These characteristics are mostly consequences of very low tectonic and volcanic activity as well as a lack of glaciation processes during the ice era owing to the low-latitudinal position of this mainland. Certainly, Australia use to be denoted as the land down under because of its low-latitudinal position within the southern hemisphere (Blewett, 2012).

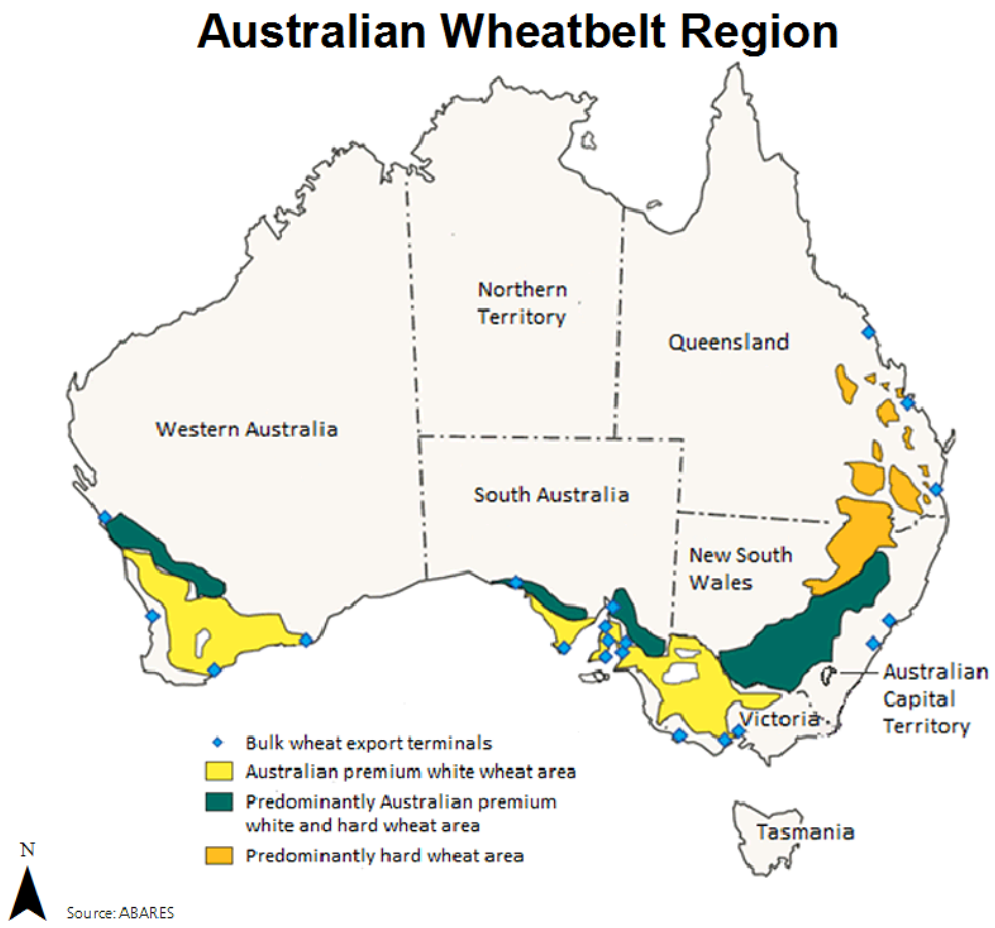


Figure 2-3. Australian Wheat-belt (also called Grain-belt) Region. This area includes most of Australia's grain agriculture. Combinations of fairly fertile soils together with a rainfall pattern which is sufficient for rainfed crop production are the main features of this area, which covers nearly 46 million hectares of Australia.

The fact that the Australian continent has been totally isolated from other land masses over the course of the last 65,000 years, has led to many unique attributes (McKenzie et al., 2004; Blewett, 2012). In conjunction with age and geological evolution, the climatic conditions have defined two others key characteristics of the Australian landscape: *flatness* and *lowness*. Dynamics of weathering and erosion processes have slowly been transforming large areas of lands into low-lying plains. It is for this reason that the average altitude found across this vast continent is a

mere 325 m with 1.4° of average slope. This is why Australia is rated as the lowest of all continents (Blewett, 2012).

These same qualities (*i.e.* geological activity, age, climatic conditions) are responsible for the low fertility of Australia's soil to the point of being recognized among the least productive but also the most different in extension and diversity by world standards. In general, Australian soils tend toward being old, salty and clayey. Although specifically in the west they tend to be sandy, acidic and even more impoverished (Blewett, 2012). Relative to the Australian standards, the youngest and more nutritionally enriched soils are usually found in the eastern parts of the continent (McKenzie et al., 2004). This is more probable due to the prevalence of more recent volcanic activity and a high influence of dust storms (Blewett, 2012).

Ironically, whilst its soils are of low fertility, Australia is at the same time one of the most biologically diverse places on the planet and representing nearly 10% of the world's biodiversity. In fact, there are significant areas in this country providing home for unique living organisms found nowhere else (Blewett, 2012; NRMMC, 2010). For example, 92% of higher plant species, 87% of mammal species, 93% of reptiles, 94% of frogs and 45% of bird species occur only in this continent (NRMMC, 2010). Unfortunately, the role and position of microbial communities related to this great biological diversity is as yet unknown. References about microorganisms inhabiting either soil or aquatic ecosystems are difficult to find or are very sparse in national/local reports or compendiums. Even though they are considered being important for the strategies of conservation of Australia's biodiversity, there are still many challenges to assess

them for being so small to be measured (McKenzie et al., 2004). Nevertheless, it is roughly estimated that to date there are 4,185 species of microorganisms described and ~ 97% remain unknown (Morton et al., 2014).

THE ENVIRONMENTS FACED BY SOIL MICROORGANISMS IN NSW

NSW comprises an area of about 809,444 km² situated in the mid-latitudes of eastern Australia (32° 9' 42" S, 147° 1' 4" E). It is bordered by the state of Queensland to the north, the state of South Australia to the west and the state of Victoria to the south. The eastern border is formed by 2,137 km of coastline which meets with the Tasman Sea (Figure 2-1). The environmental conditions within these borders shape the habitats in which has evolved the microbial populations we found living into the soils of NSW. For this reason, it is necessary to start outlining the status of these environmental conditions today, which is exposed in detail later in this chapter when describing the physiography, climate, soils, landscapes and biodiversity attributes of this terrain.

Briefly, NSW is characterized by a diversity of landscapes - Aeolian, Erosional, Fluvial and Coastal types – under highly variable but moderate climate (NARCliM, 2014). NSW is divided from west to east into three geographical sections: *The Western Plateau, Interior Lowlands* (*Western Plains) and Eastern Uplands* (including the *Great Dividing Range*) (Figure 2-4). *The Western Plateau* is a

miscellaneous collection of mountains, plains and dunes that are sparsely vegetated with grassland and desert ecosystems under either a semi-arid or an arid climate. *Interior Lowlands* or *Western Plains* are extensive low-flat areas under arid or semi-arid conditions. These plains cover nearly two-thirds of the state; however, the human population found in this area is very sparse when compared to that of coastal regions. Heading towards the east, the Western Plains begin to merge with those of the agricultural plains. This zone represents the most important area for the national grain production and the primary economic activity of rural NSW. The agricultural plain gently disappears to the east in an undulating landscape that finally meets with the slopes, tablelands and mountainous section of the *Eastern Upland*. The *Eastern Upland* is a massive elevated strip of land that includes both the Great Diving Range as well as the Great Escarpment. The *Eastern Upland* extends from north to south and reaches average peaks of no more than 1,000 m. To the east side, *Eastern Uplands* slope down toward the coastal areas that are the most populated zones of NSW. The major climatic variations across the state occur from the northeast to the south. Summer rainfalls and relatively dry winters characterize northeast side. However, winter rainfall under cold conditions is the more common pattern in the south (NARCliM, 2014).

Regarding NSW soils, *Vertosols*, *Kandosols*, *Calcarosols* and *Sodosols* dominate large areas of the western region. To the east of the Great Dividing Range, an increase can be found in the pedodiversity with a mixture of smaller areas comprising of approximately 12 different soil types (*ASC in* Figure 2-37). In general terms, the most important complication affecting NSW soils is degradation in which acidification, salinity and wind erosion signify one of the most difficult environmental management for the state (EPA, 2012). The generally well-nourished and

comparatively younger soils are commonly found on the eastern side of the Great Dividing Range towards the coastal zones. However, few other fertile zones can still be found to the west of the Great Dividing Range towards the agricultural plains zone, encompassing the *Wheatbelt Region* (McKenzie et al., 2004).

NSW's biological diversity contributes significantly to Australia's biodiversity. Example of this contribution can be seen within the 18 different bioregions found across NSW, and furthermore, two of them have been designated as among Australia's 15 National Biodiversity Hotspots. Additionally, these biological resources are part of the protected areas of reserve ecosystems, which nowadays correspond to approximately 8.8% of NSW terrain (DSEWPaC, 2012a; EPA, 2012).

Regarding to soil biodiversity and specifically to soil micro-biota, as it was commented above when describing Australia's biodiversity, there is a lack of outright information related to the current status of these tiny organisms also within NSW biodiversity overview. The Biodiversity Strategy for NSW – on behalf of a major governmental plan for the protection and conservation of Australia's natural resources - recognizes microorganisms as a source of genetic diversity and life forms that play a significant role on providing ecosystems functions and ecosystems services (NRMMC, 2010). Even though it is known that their diversity is declining, there are not official estimations, for example, about the total number of species and/or taxonomic groups composing these microbial communities (NRMMC, 2010). For these reasons, the identification of Australia's microorganisms, particularly those living into soil environments, is considered as a

highest priority research in Australia (Australian Government, 2001).

In sum, the most significant environmental attributes shaping NSW define the resulting soils and, in consequence, the *habitat* of the microbial communities living there. Main of these NSW attributes are detailed below based on a brief description of physiography, climate, soils and landscapes accompanied with different maps that support this and other supplementary environmental information. In addition, a globalised view of the previous attributes and others contributing factors (*e.g.* lithology, landform, flora, fauna) are presented based upon a landscape approach *via* the NSW bioregions from the Interim Biogeographical Regionalization (IBRA7) (DSEWPaC, 2012a). This IBRA7 is a classification system that provides a useful ecological view of the environmental gradients faced by the microbial communities.

PHYSIOGRAPHY OF NSW

There are quite different systems of Australia's regionalization (formal and informal) upon the different politico-administrative and/or socio-cultural purposes, *e.g.* geography, public services, etc. Physiography is considered the basis of a regional division for understanding the characteristics of natural resources that are strongly dependent on both geographical positions and landforms; for example, distribution of soils and natural vegetation. A physiographic description provides general references of geomorphological characteristics grouped on the basis of a physical geographical view *i.e.* landforms (ACLEP, 2011).

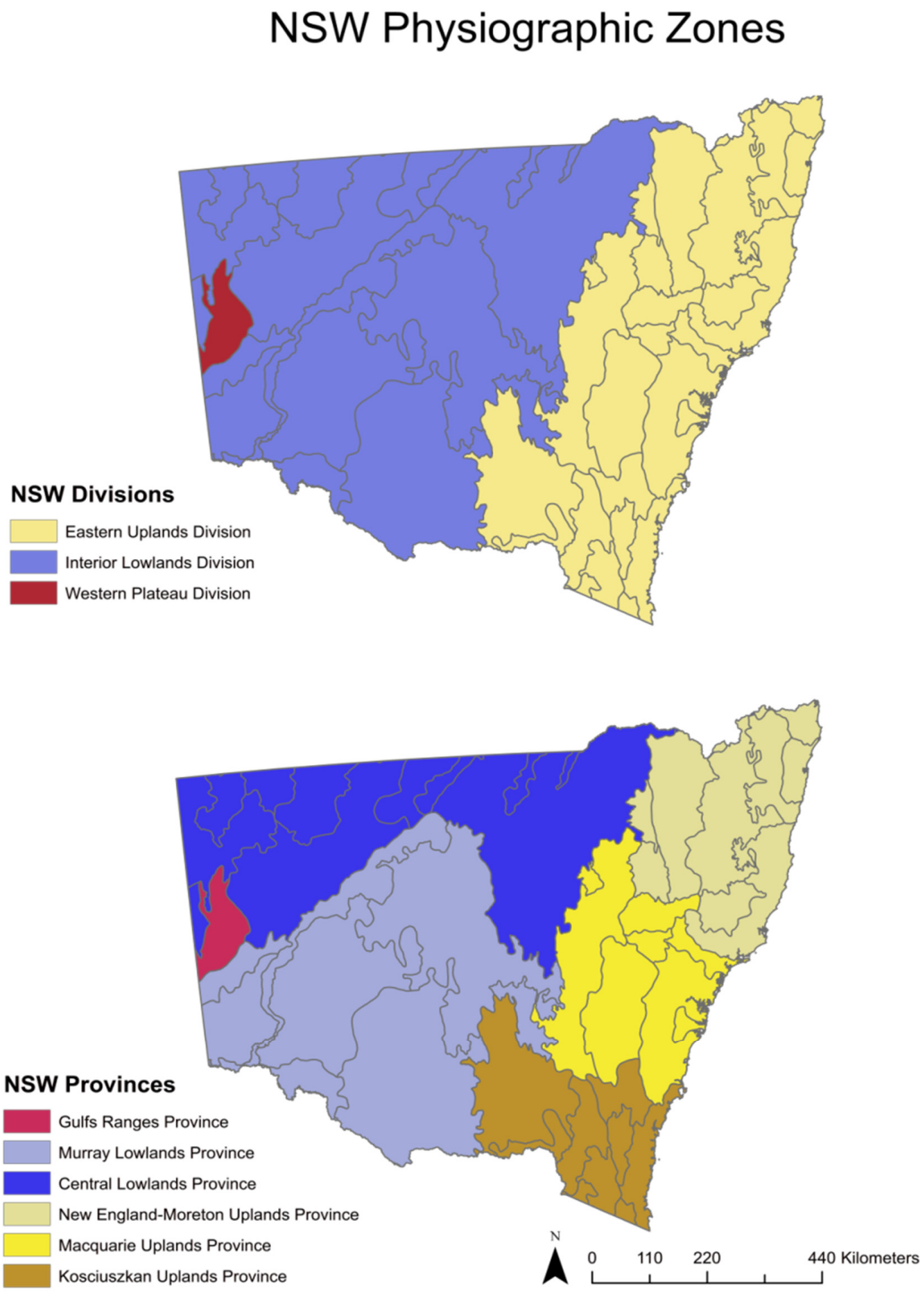


Figure 2-4. Principal natural physiographic features dividing NSW (ABARES, 2014).

Basically, NSW mainlands and island areas occupy 800,642 km² of land partitioned in three physiographic divisions (Figure 2-4). Described from east to west these divisions are named: *Western Plateau Division, Interior Lowlands Division and Eastern Uplands Division* (Blewett, 2012; McKenzie et al., 2004). These divisions are mainly different in term of their geological nature and landforms (slope and relief). Following the hierarchical aggrupation, each division is classified into provinces and regions[vi]. Provinces share similarities in terms of slope and relief but also in terms of soil orders and water balance. As a complement and specifying more details, the regional clarification involves the identification of soil suborders, ages of land surfaces and regolith materials (ACLEP, 2011).

(i) Western Plateau Division

Bordering the state of South Australia, *The Western Plateau Division is made up of a* mosaic of plains, plateau, ranges and dune fields. The only province in this division is called *The Gulf Ranges*, characterized by the presence of The Adelaide Hills and lowland areas which extend until Broken Hill. The area covers nearly 12,000 km² and is positioned about 60 - 200 m above sea level (Hill, 2004). The only region is the *Barrier Ranges Region*, which is formed by a range of hills and mountains within an area of mostly undulating lowlands of granite and metamorphic. The range roughly trends from north to south and in some areas rises up to 473 m. The landscapes expose mostly moderate weathered bedrock (> 50%) and soil on bedrock (20 - 50%) (Hill, 2004; NARCliM, 2014).

[vi] *Division: defined upon broad physiography (slope and relief) and geological attributes. Map scale 1:10 million.*
Province: defined upon physiography, water balance, dominant soil order and substrate. Map scale 1:2.5 million.
Region: defined upon physiography, regolith materials, age of land surface, water balance, and dominant soil suborder. Map scale 1:1 million (ACLEP, 2011).

(ii) Interior Lowlands Division

The Interior Lowlands Division also referred as *The Western Plains*, corresponds to younger basins of vast flat riverine sediment which has been deposited by the ancestral streams of the Murray-Murrumbidgee and Darling rivers (Blewett, 2012). This portion is separated into two provinces: *Central Lowlands Province and Murray Lowlands Province*.

Firstly, *Central Lowlands Province* covers about 217,000 km^2 in the north and north-west of NSW. It is characterized by an internally draining interior lowlands area showing very highly weathered bedrock (>50%). Secondly, *Murray Lowlands Province* is located in the south and south-west of NSW and, represents approximately 260,000 km^2. It is basically composed of flat alluvium (> 50%) and is rather similar to the Murray sedimentary basin (ABARES, 2014).

(iii) Eastern Uplands Division

The *Eastern Uplands Division* uplifts on the further most eastern side of NSW. This division corresponds to approximately 38% of the land surface of NSW and stretches from Queensland in the north to Victoria in the south along the eastern seaboard. North to south the landform patterns of this division reveals structural contrasts in a general grain of relief, based on which are recognized the three different provinces found here: *New England-Moreton Uplands Province*, *Macquarie Uplands Province* and *Kosciuszkan Uplands Province* (ABARES, 2014). Following the eastern seaboard from north to south, the first province situated in the north, sharing the boundary of Queensland State, is the *New England-Moreton Uplands Province* (*Northern Tableland*

Region[vii] *included*).

New England-Moreton Uplands Province covers 100 km^2 and is characterized by higher uplands descending toward the eastern seaboard. This particular province encompasses a wide range of landforms and geological substrate typified by an abundance of weathered bedrock (> 50%) and soil on bedrock (20 - 50%). From east to west this substrate and land surfaces ranging from ridges, valleys, plains or hills derived of metamorphic, volcanic or alluvium materials; to plateaus and dissected plateaus composed by metamorphic and basic intrusive rocks and undulating granitic and basaltic materials. Bordering the coastline we find a series of coastal lowlands of weak sedimentary rocks, with littoral and alluvial plains and plateaus dissected into narrow strike ridges and valleys (ABARES, 2014).

Towards to south and, relatively, in the middle of the eastern seaboard, take place *The Macquarie Uplands Province* (in which the *Central Tablelands* [viii] is included). This zone corresponds to approximately 100 km^2 which is mostly conformed of dissected plateaus on sub-horizontal resistant sandstones, predominantly from the *Sydney Basin*. The eastern part of this province is characterized by tablelands stepping down to west and breaking into detached hills. To the east, the middle zone uplifts tablelands of granitic and basaltic substrate and minor lowlands; also it includes a dissected volcanic pile from *Mount Canobolas volcano*[ix]. The coastal section mostly contains deeply dissected sandstone plateau (ABARES, 2014).

[vii] *Northern Tableland Region is a regional division according to an informal regionalization of NSW upon a relative position to Sydney.*
[viii] *Central Tablelands is a regional division according to an informal regionalization of NSW upon a relative position to Sydney.*
[ix] *Mount Canobolas is an extinct volcano located in the Central Tablelands of NSW.*

Finally, the southern section of the *Eastern Uplands Division* is occupied by the *Kosciuszkan Uplands Province*. This province borders with the Victoria and expands to approximately 86,550 km^2. It is mainly characterized by mountains and plateau ranging from the highest points in Australia to the coast. Into this range of landforms, this province exhibits more than 50% of regolith materials of the type of saprock[x]. In the west side, bordering the Interior Lowland Division, this zone is characterized by ridges and minor tablelands that slope down westwards and break up into detached hills with intervening alluvial valley floors. Towards the east, this landform gradually changes to dissected high uplands with some periglacials features; to plains with separating strike-aligned hills, which occasionally reveals closed lake basins in between; and to undulating upland plains with some tabular basalt relief and granite tors. In the coastal zone, eastward of the Great Escarpment Figure 2-5, the landscape converts into deeply dissected steeply sloping plateau margin in metamorphic and granite materials.

Topographically, seen from west to east, the *Eastern Uplands Division* is very well defined into three different sections: *The Western Slopes*, *The Eastern Highlands* (commonly known as the *Great Dividing Range*) and *The Great Escarpment*. *The Western Slopes* is generally an undulating and transitional terrain merging to the west with the extended plains of the *Interior Lowlands or Western Plains*. This zone has been developed the most fertile agriculture in NSW and it was the first zone explored by this research ($NS_{\text{-transect}}$). Geomorphologically, the western slopes can be seen as similar to a ramp connecting the *Western Plains* with the uplifted *Eastern Highlands* (ACLEP, 2011; Blewett, 2012; Hill, 2004; McKenzie et al., 2004).

[x] *Saprock: saprolith -earth's regolith- material caracterized by fractured bedrock with weathering restricted to fracture margins.*

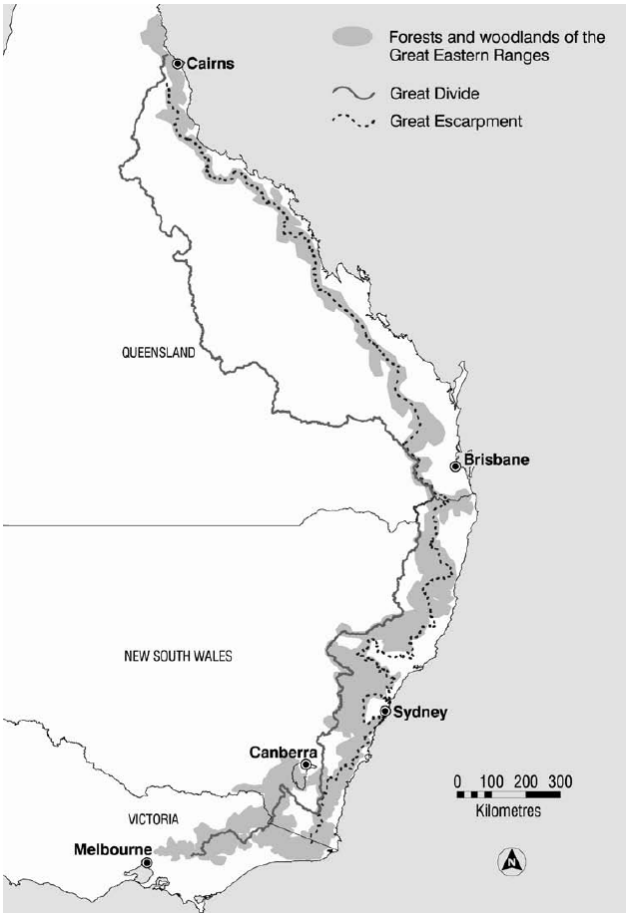


Figure 2-5. Australia's Great Eastern Ranges: Great Dividing Range and Great Escarpment.

The *Eastern Highlands* or *Great Dividing Range* is a zone of elevated tablelands with gently undulating terrains and broad plains which also concentrate the highest lands of Australia along their south most part. At the east side of this Great Divide, where the higher tablelands areas begin to slope down towards the coast, uplifts The Great Escarpment (Figure 2-5).

This escarpment is characterized by long and vastly forested cliffs meeting with elevated areas rising above several hundred metres. These features delineate a boundary between area of low geomorphological dynamism found on the inland side, in contrast, to those zones of more active processes found on the coast side (ACLEP, 2011; Blewett, 2012; Hill, 2004; McKenzie et al., 2004).

CLIMATE OF NSW

The climate of NSW is highly variable and is generally considered as a temperate zone (NARCliM, 2014). The largest climatic variations depend upon one's proximity to certain geographic features; such as the east coast, the mountainous area of the Eastern Highlands (Great Dividing Range) and the desert region towards the north-west (Figure 2-7). For example, the coastal areas are greatly influenced by the warm waters of the Tasman Sea. As result, moderate temperatures dominate together with a high level of moisture, creating conditions for abundant rainfall. The Great Divide Range also receives considerable amounts of precipitation carried west from the coastal regions by onshore winds. This range also acts simultaneously as a massive barrier between coastal zones with those lands on the interior side of the range, reducing considerably

the rainfall patterns of western NSW.

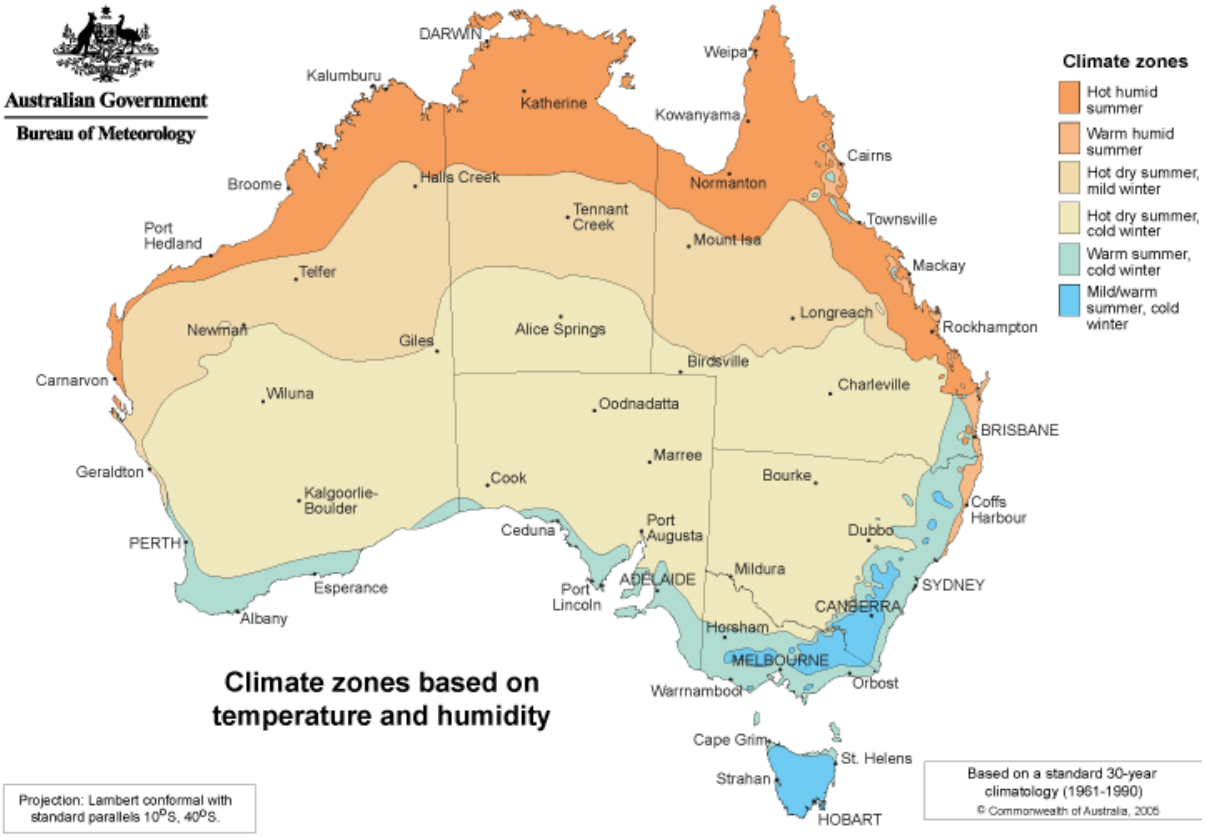


Figure 2–6. Climate zones based on temperature and humidity [xi] (BOM), 2005).

In Australia, there are three climate zonification schemes based on three different methods of classification to differentiate the climatic variations in the mainland territory: *temperature/humidity*, *seasonal rainfall* and *vegetation* (*Köppen* classification). According to the first classification scheme, NSW expose four of the six of the Australia's climate zones upon

[xi] *The temperature and humidity zones map shows the climate of Australia classified according to temperature and humidity properties across the country.*

temperatures and humidity. These zones oscillate from *hot dry summer/cold winter* in the west, to *mild-warm summer/cold winter* in the east, changing to *warm humid summer* toward the north-eastern coast (Figure 2-6).

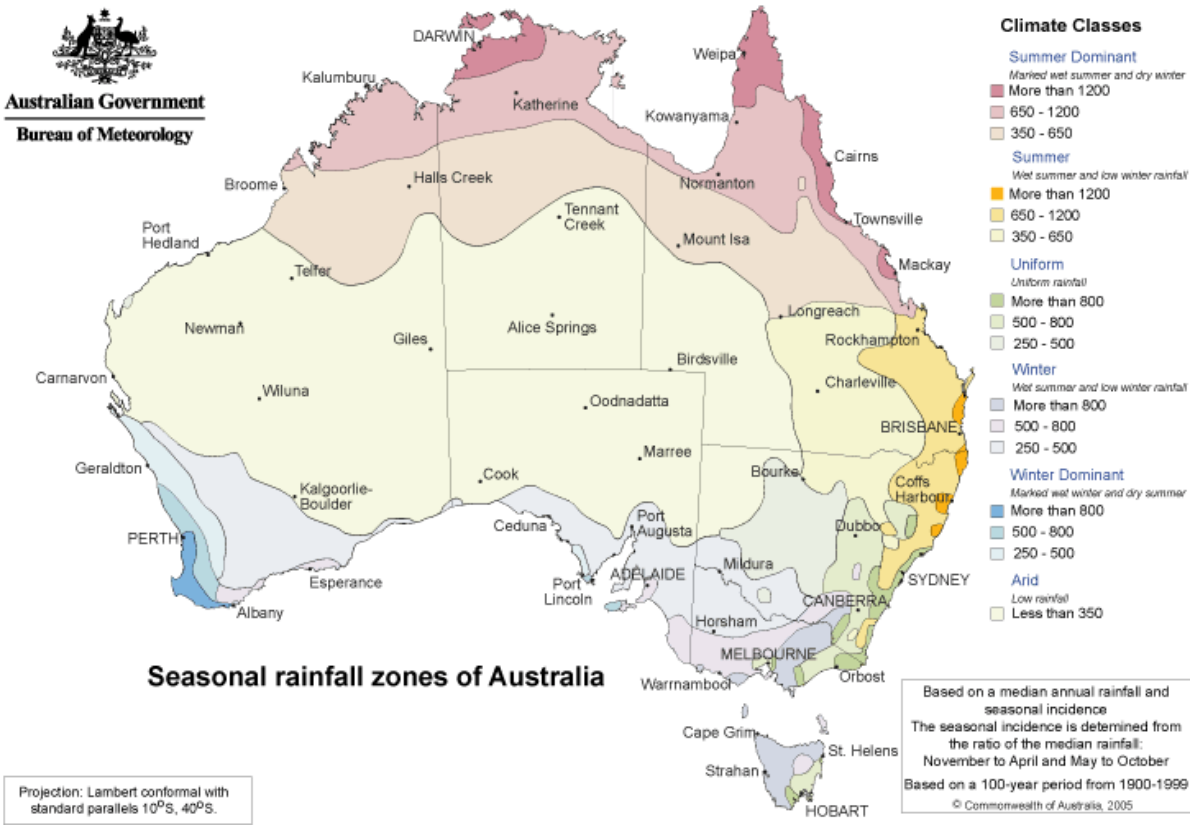


Figure 2–7. Seasonal rainfall map of Australia[xii] showing the major climatic zones defined based on the median annual rainfall and seasonal incidence (BOM), 2005).

The second zonification scheme classified the seasonal variation of the precipitation pattern is into seasonal rainfall zones as it is shown in Figure 2-7. In this case, NSW encloses four of the

[xii] *The seasonal rainfall maps use the differences between summer and winter rainfall across Australia to identify six major climate zones.*

six Australia's climate classes, by which most of the central part of the state (mostly encompassing the *Interior Lowlands* Division) exposes *uniform class* (referencing uniform rainfall between summer and winter). Conversely, the main drastic variation is giving from west to east, particularly in the north territory where the zones varies from *arid class* (low rainfall) characterizing the central and north zones of the western area; to the *summer class* (wet summer and low winter rainfall) encompassing the central and north zones of the eastern area. In the southern NSW, instead, these classes range from *winter class* (wet winter and low summer rainfall) in the western area, to *uniform class* (uniform rainfall) towards the eastern coast.

The third classification system on the basis of the Köppen vegetation scheme, classified the Australia's climatic zones into six major groups: *Equatorial, Tropical, Subtropical, Desert, Grassland and Temperate*. NSW possesses four of them, which in most of the NSW territory from west to east are found in order the *Desert*, *Grassland* and *Temperate* groups. The *Subtropical group* is only found in the northern east side of NSW, excluding the elevated zones of the *Eastern Uplands Division* categorized as a *Temperate zone*.

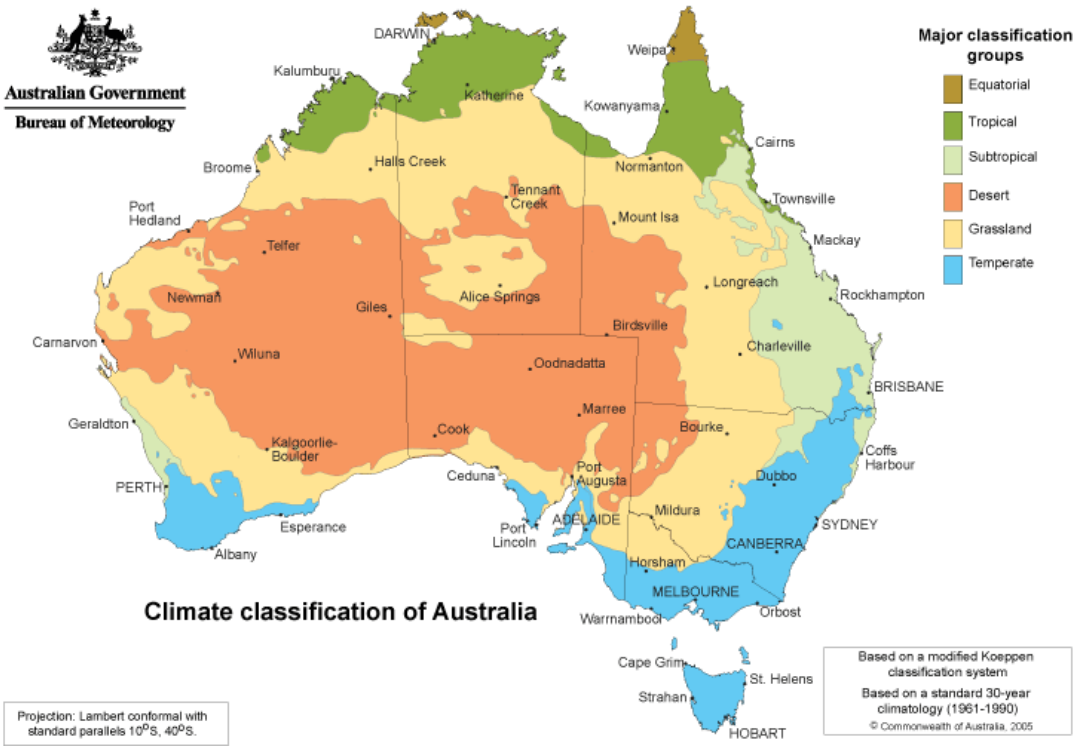


Figure 2−8 . Climate classification map of Australia defined based on the climatic limitation for different types of native vegetation (Köppen methods modified) (BOM), 2005).

(i) Temperatures

The mean annual NSW temperature ranges between 3°C and 21°C (Figure 2-36). This fluctuation mainly occurs in a diagonal approach, from the very warm far north-west (Strzelecki Desert Plains) to the very cold southern alpine regions (Snowy Mountains). Average seasonal temperatures in the north-west vary from 31-36 °C in summer to 6-11 °C in winter. Whereas in the area of the south-east including the Snowy Mountains, average temperatures range from 14 to 20 °C in the summer and may go down to -5 to 4 °C during the winter (ABARES, 2014; NARCliM, 2014).

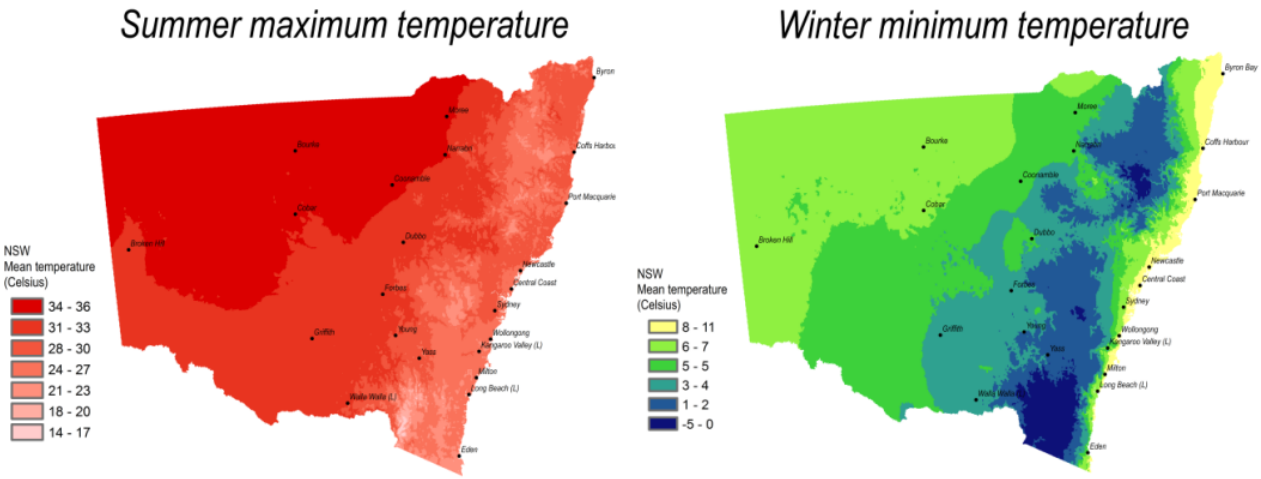


Figure 2–9. NSW extremes temperatures during summer and winter. Temperatures values correspond, respectively, to the average daily maximum and minimum extremes.

Future projections have been informed by the Overview of NSW Climate Change Report 2012, estimating that all the temperatures variables (average, minimum, maximum, etc.) will have an increase across all of the state with the north-west area the greatest affected. These increases could be approximately 0.7 °C in the near future and up to 2.1 °C in the far future (NARCliM, 2014).

(ii) Rainfall

In general terms, the precipitation pattern for NSW is highly variable. The most remarkable phenomenon is the drastic decrease in rainfall from east to west; once we get away from the Eastern Mountains. Another important characteristic is the seasonal differentiation found across the state. While the north-east has an abundant summer rainfall and a relatively dry winter, the southern part of NSW experience higher volumes of winter rainfall under cold

conditions. There are also sporadic rainfall events over the arid north-west at any time of the year - but they are more likely in summer- (NARCliM, 2014).

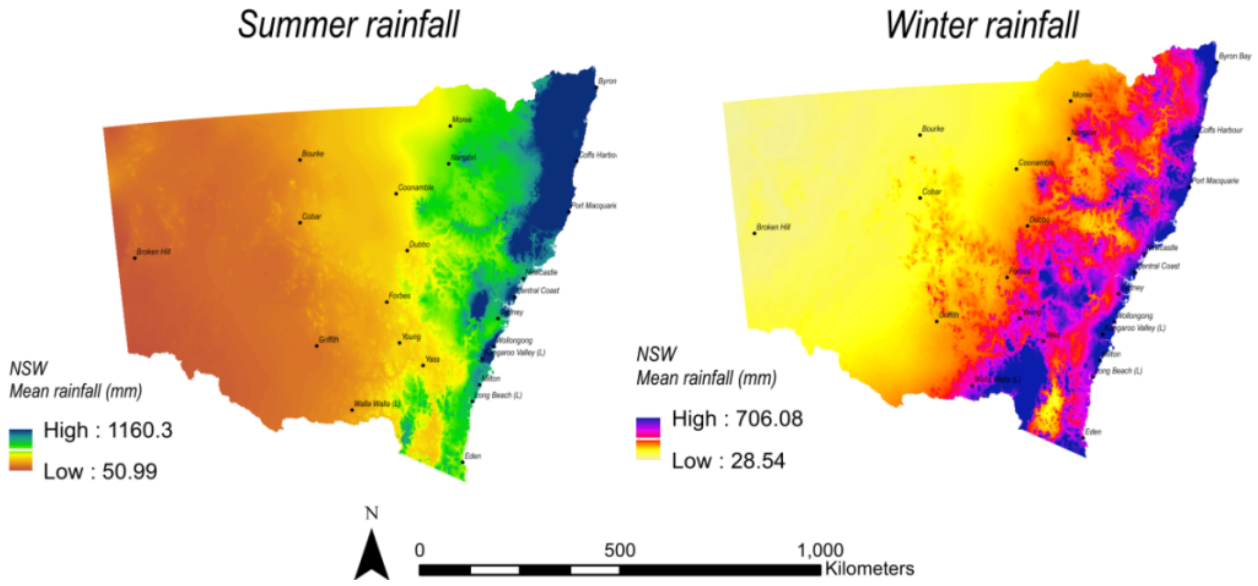


Figure 2−10. Summer and winter mean rainfall pattern across NSW.

The mean annual rainfall across the state is 1,138mm. This average ranges from 1,300 to 3,200 mm per year (concentrated on the far east-north coast) to 170 to 370 mm per year (mostly concentrated in the far western NSW) (ABARES, 2014). It is projected that by 2,030 (near future) and 2,070 (far future) in NSW there will be a decrease in the mean spring rainfall and an increase in autumn's rainfall (NARCliM, 2014). Increases in autumn's rainfall will be given all across the state. However, spring rainfall will be varying independently across some regions. Specifically, spring rainfall will decrease mostly for inlands regions and southern NSW. Conversely, the north coast between Queensland border and Newcastle will experience increases in spring rainfall.

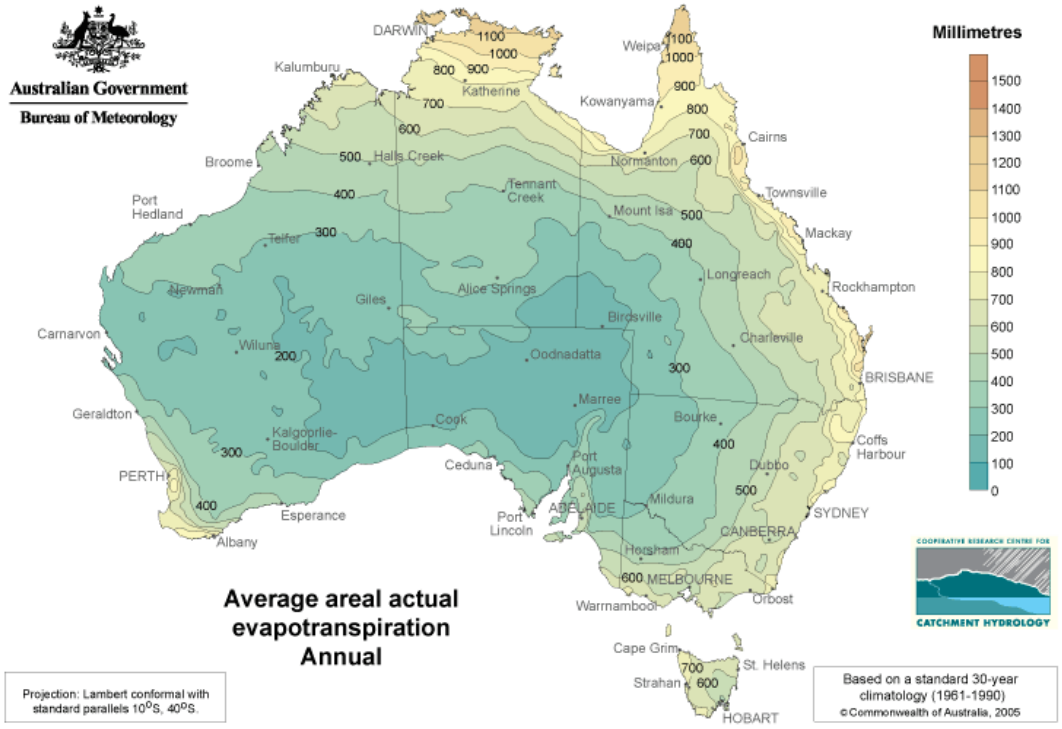


Figure 2–11. Average annual evapotranspiration distributed across Australia (based on areal actual calculations[xiii]).

(iii) Evapotranspiration

Evapotranspiration (ET) from both vegetated and un-vegetated land surfaces is affected by climate conditions, availability of water and vegetation patterns. On the basis of the calculations of the areal actual ET across NSW, this variable increases gradually west to east, from being very low ranging 200-300 mm/year in the far western zones, until it reaches series of about 800-900 mm/year in the northern coastline.

[xiii] *Areal actual ET is the ET that actually takes place, under the condition of existing water supply, from an area so large that the effects of any upwind boundary transitions are negligible and local variations are integrated to an areal average. For example, this represents the evapotranspiration which would occur over a large area of land under existing (mean) rainfall conditions (BOM, 2005).*

FROM LANDSCAPES TO SOIL ECOSYSTEMS

(i) Types of soils across NSW

The soils of NSW are highly diverse and widely extended; tending to be old, salty, clayey (Blewett, 2012; Charman and Murphy, 1991). By Australian soil fertility standards, NSW has comparatively the youngest soils and landscapes encompassing a quite fertile area (McKenzie et al., 2004). This fertility can be in part attributed to the Australia's formation process that occurred from a westerly to easterly direction. In geological terms, this means that NSW was one of the last portions of land to become defined in early Australia (Blewett, 2012; McKenzie et al., 2004).

It has been recognized that other causes for poor soil fertility in Australia arose from a lack of nourishment from volcanic activity together with deep weathering processes combined with low relief. Nonetheless, NSW when compared with other regions presents the most recent volcanic events as well as still some traces of the existence of icebergs from the ice age. In addition, NSW does not present extended areas of deep weathering processes, except few areas in the far north-west. In fact, NSW most commonly shows minor weathering and moderate to high relief (Blewett, 2012).

Soils and landscapes distribution are clearly delineated by the physiographic features of NSW, based on which, the main differentiation is given between *Interior Lowlands* and *Eastern Uplands Division* (ABARES, 2014). These natural features, undoubtedly, execute a significant effect on the

soil formation process when influencing each of the soil formation factors – organism, topography, climate, parent material and time. For example, Figure 2-12 relates the distribution of lithological material and soil types distributed across NSW. It is clear how the regolith materials – which cover nearly 80% of Australian surface - remain limited to the Western Plateau as well as to the Interior Lowlands. Whereas, conversely the *Eastern Uplands* exhibit mostly sedimentary materials (ABARES, 2014; Blewett, 2012).

The Australian Soil Classification has distinguished twelve soil orders in NSW, where *Vertosols* are the most widely extended (Figure 2-13). Even though *Vertosols* are present almost across of all of the state, they are mostly in the transitional zone from the *Inland Lowland Division* to the east - from undulating terrains toward the upper surfaces at the beginning of the *Eastern Upland Division*. In terms of extent, *Vertosols* are followed by *Calcarosols* and *Sodosols* which dominate in the riverine plains of south-central and southwestern of NSW. Then, these are followed by *Kandosols* and *Rudosols*, which are largely extended in the north and north-west of the state. The *Eastern Uplands Division* has an increase in pedodiversity and gives place to more soils types when compared with western NSW. However, these soils orders involve considerable smaller areas. The soil orders represented in the eastern side of NSW ranges among *Chromosols, Kandosols, Sodosols, Dermosols, Tenosols, Ferrosols, Kurosols, Podosols* and *Hydrosols* (ABARES, 2014).

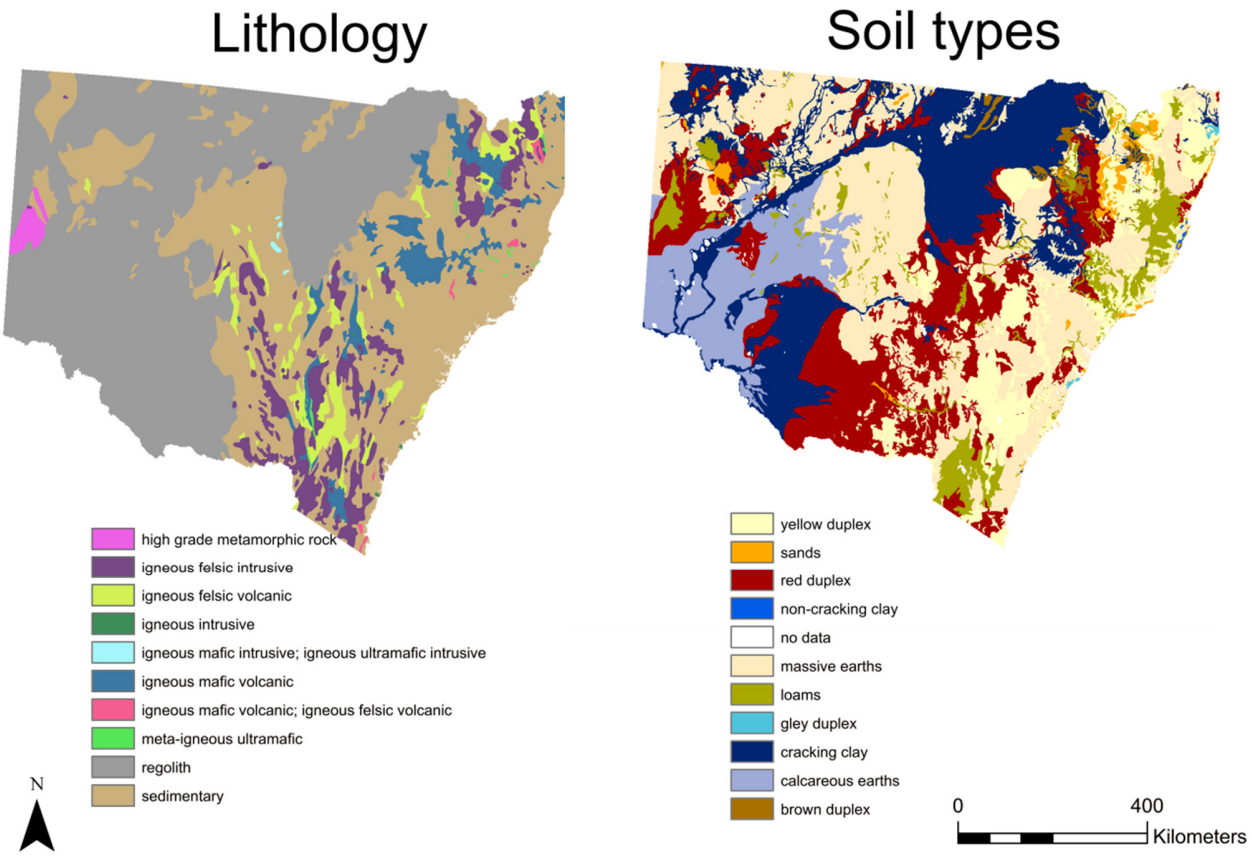


Figure 2–12. Lithology[xiv] and soils types[xv] distributed across NSW (ABARES, 2014)

[xiv] *Major lithological types based 1:2.5million geology map.*

[xv] *Soil types catetegorized based on The Factual Key of Northcote scheme (1979). This is a soil classificaton systems that uses field observable soil morphological data. It has been widely used in Australia during the last 30 years and most notably formed the basis for characterising soils in the Atlas of Australian Soils (Mckenzie et al., 2000).*

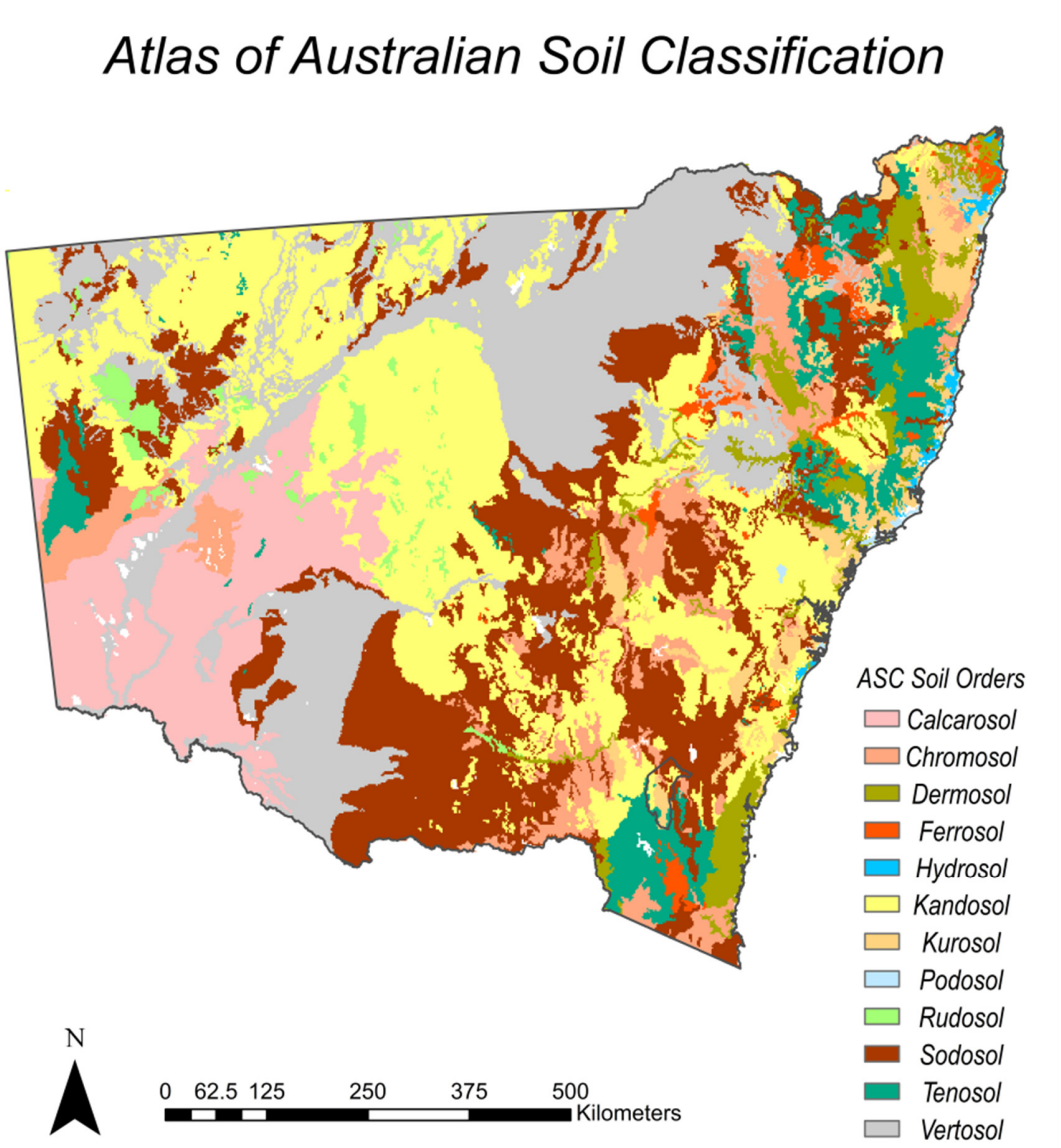


Figure 2−13. The Atlas of Australian Soils Classification[xvi] differentiates12 types of soils distributed across NSW (CSIRO, 2001).

[xvi] *A table that converts the Atlas of Australian Soils mapping units to an Australian Soil Classification soil order was compiled to aid the production of concepts and rationale of the Australian Soil Classification (1997). This map was extracted from the Australian spatial layers 2014 datapack provides by Australian Bureau of Agricultural and Resource Economics and Sciences* (ABARES, 2014).

(ii) The most typical landscapes building up NSW soils

Pedodiversity is highly dependent on landscapes diversity; in which, NSW is considered vastly enriched. Therefore, landscapes at determined locations give an essential contribution to distinguish the mentioned soil orders distributed across NSW. In this area of Australia, most of the landscapes were formed as result of specific climatic events (*e.g.* long periods of drought) in conjunction with weathering, erosional and depositional processes (Blewett, 2012; McKenzie et al., 2004). As consequence, these landscapes today mainly range among *Aeolian, Coastal, Erosional* and *Fluvial* types (McKenzie et al., 2004). Each of these landscapes are briefly described emphasizing the context of NSW soil formation.

a) Aeolian landscapes: *e.g.* continental dunefields and dust mantles in arid zones

Aeolian landscapes are mostly part of arid and semi-arid zones and can be exemplify by continental dunefields and dust mantles landscapes. Continental dunefields - distinguished from coastal dunes – are found to the central-west border of NSW. The dunes landscapes characterized by a simple development of sandy soil profiles of deep red colour, are largely product of the dominant presence of iron oxide (hematite). Instead, dust mantles can be found either as a distinct mantle or as a component of the soil profile. Soils derived from dust deposits - also known as *parna*[xvii] grounds - with well-drained condition such as, *Kandosols* and *Dermosols*, sustain many of the best NSW cropping ecosystems located in the agricultural plains of the *Interior Lowland Division*. These soils are usually very well structured, permeable, aerated and of an outstanding water storage capacity (Figure 2-37). Further east in wetter locations, it is more usual to see dust incorporated into the soil profile. For example, it has been reported deposition

[xvii] *Parna is an Australian aboriginal word which means sandy and dusty ground* (McKenzie et al., 2004).

of dust from western dunes into soil profiles of forested areas in the western side of the Snowy Mountains (Blewett, 2012; McKenzie et al., 2004).

b) Fluvial landscapes: east-coastal rivers and riverine plains

Fluvial landscapes compose other typical environment in the humid areas of southeaster NSW; and likely, the most archetypal examples within these landscapes are both east-coastal rivers and riverine plains. East-coastal rivers landscapes are product of multiples conjugations of rivers, water flows, streams and sediments from different sources. These factors, in conjunction, have led to form a wide variety of soils. The development status of these soils is also variable upon the fluvial sequences from young stream banks to well-developed terraces. Stream banks soils are shortly differentiated to simple profiles and they are found frequently flooded; instead, soils on terraces are strongly differentiated, specially between A and B horizons, and they are no longer flooded, although, in many cases exposed to erosion (Blewett, 2012; McKenzie et al., 2004).

Riverine plains landscapes also play an important role by shaping NSW soil genesis. In NSW, there are two large riverine plains: the Darling Riverine Plain and the Riverine Plain of south-eastern Australia; both originated from the Murray-Darling basin (1,072,000 km^2). Murray-Darling is the Australia's longest river network at over 3,670 km in length extended from South Australia State to Queensland State and, by which, covers 14% of Australia's surface (Figure 2-14) (ABS, 2015a; Blewett, 2012; McKenzie et al., 2004). Perhaps, this elongated extent becomes even more significant when it is considered the distinctive low water flows and, subsequently, low gradient of this river system; in fact, the basin does not discharge more than

767 m^3/sec[xviii]. In effect, this low-gradient signify an important fact influencing soil formation in the area cover by this catchment (ABS, 2015a; Blewett, 2012; McKenzie et al., 2004).

For instance, one of the main factors differentiating soils of these landscapes is the sedimentation processes. Sedimentation occurs in upper catchments dominated by hillslopes terraces, where the deposition of coarse sediments take place near to the banks of the stream but the finer sediments are deposited further away because of the gentler water flow. As result, most of the soils found in these areas, range from coarse textured on levee banks to Vertosols on clay plains. Riverine landscape is another common example composed by soils formed upon sedimentation.

c) Erosional landscapes: *e.g.* ravine and arboreal cliffs of the Eastern Uplands

Erosional landscapes are notorious features of eastern NSW. These kinds of landscapes arise when erosional processes, such as wind or water, disrupt the land surface removing soil, rocks or dissolving materials. For example, typical cases are surfaces exposed with reduced vegetation cover, lowering of riverine streams and depositional lands (*e.g.* lava, dust). Moreover, when this erosion affects uplifted land masses derived from tectonic activity; such is the case of the *Eastern Upland* in NSW. For example, gullied lands of the Southern Tableland in the Great Divide and the forested lands of the Great Escarpment are representative cases of erosional landscapes. The gully areas are mostly part of either sloping or flat areas derived from deep deposits or alluvium where the vegetation mantle has been abruptly disturbed most likely through fire or grazing.

[xviii] *In terms of comparison, Mississippi-Missouri basin in USA reachs a discharge of about 18,400 m^3/sec, occupying an area of 3,202,230 km^2 by its river system, which scopes approximately 6,275 km in lengh* (Blewett, 2012).

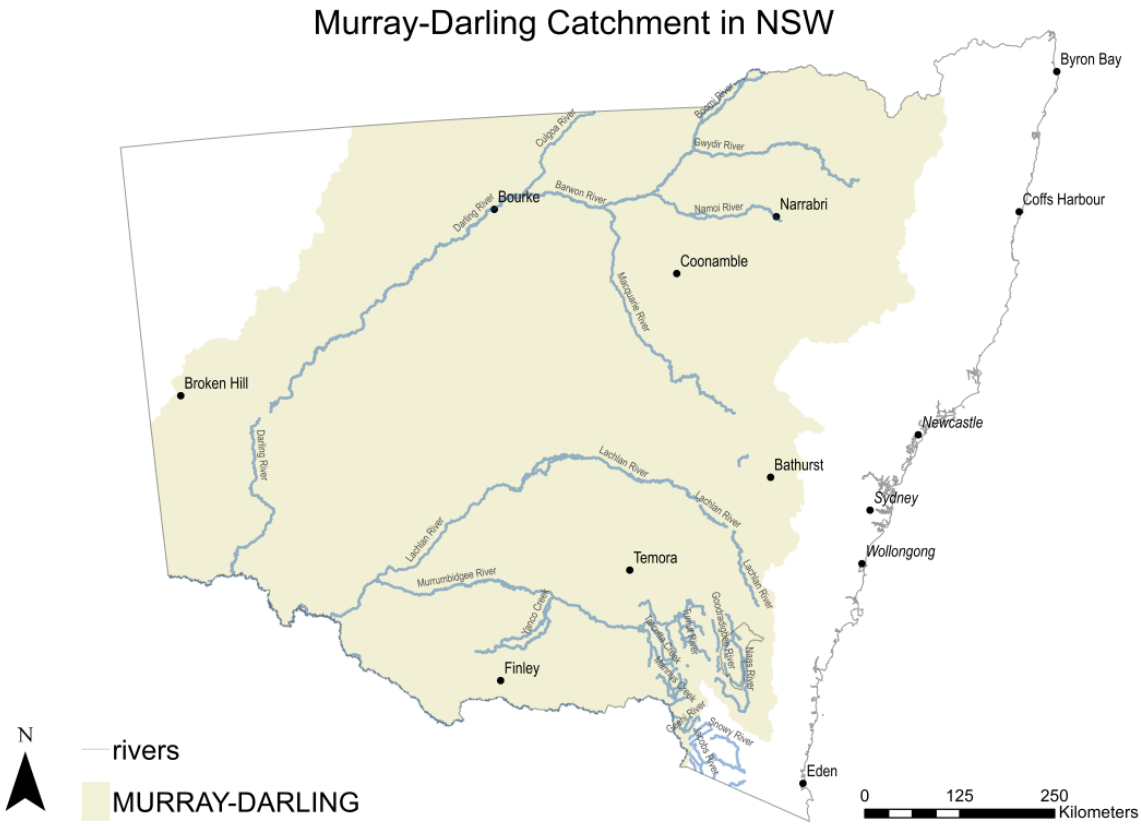


Figure 2–14. Murray–Darling Catchment covers 75% of NSW land area, in other words, 14% of the Australian continent. The Basin contains Australia's three longest rivers, the Darling (2,740 km), Murray (2,530 km) and Murrumbidgee (1,690 km). Significant proportions of the Basin's area are comprised of agricultural land (67%) and native forest (32%) (ABS, 2015a).

On its own, the Great escarpment, which is the greater geomorphic boundary in NSW, is one of the best evidence of landscapes formed by erosion. Firstly, this sharp landmass uplifted together with the *Eastern Uplands* and subsequently, it was erosional processes by rivers enhanced its abrupt relief. This is a very dynamic ecosystem where it is difficult to preserves old soils and landforms, however, an enriching rainfall pattern and proximity to coastal areas have allows to establish extensive forested areas. Forests generate a rich bio-mantle contributing to soil fertility and contrasting soil textures. This areas support a diverse type of soils; shallow, gravely and

relatively young soils (*Rudosols*) are common in the steeper slopes and sharply zones; conversely, more fertile and deeper soils are found in lower slopes and gullies deposits areas.

(iii) Principal soil conditioning factors in NSW

Certainly, all these NSW landscapes are characterized by their soils and *vice-versa;* thus, the nature and characteristics of NSW soils are intimately linked to the landscapes in which are contained (McKenzie et al., 2004). This is why we also find within certain features, the main causes of pressure upon soil ecosystems. In global terms, the most common stresses affecting NSW soils – as most of Australia's soils – are, acidification, salinity and erosion; all of them aggravated when either there is an invasion of non-native vegetation or a loss of native vegetation (Blewett, 2012; EPA, 2012).

For instance, there are many natural sources of salts in Australian landscapes (*e.g.* parent material, groundwater, coastal marine sediments, dry lake beds, etc.); however, salt accumulation and subsequently soil salinization can be promoted and/or intensified by certain attributes of a given landscape such as longevity, landform and vegetation (Blewett, 2012; McKenzie et al., 2004). In NSW, one common case is the exacerbation of salt accumulation upon the low-gradient and slow water moving of the river systems – as a consequence of the dryness and flatness. This phenomenon is usually found in arid zones or even afterwards a long drought period. For example, the riverine plains of the Murray–Darling Basin being one of the lowest parts of the Australian continent - detailed above when describing NSW fluvial landscapes - use to expose serious problems of salinization (Blewett, 2012). Even more, also in wetter zones,

towards the south-eastern part of the Murray–Darling Basin, salinity can be intensified as result of invasive vegetation along way of the river streams. For example, the presence of *Salix* species or Willow trees, which over and above of being consuming large volume of water, tend to grow up on the stream bed impeding the flow ability of the water and increasing the amount of salt accumulated (Blewett, 2012; EPA, 2012; MDBC, 1999).

Another common phenomenon is call dryland salinity, which is the accumulation of salts on the soil surface in un-irrigated zones. The main causes of this phenomenon also rely on the combination of old landscapes with low relief, but in this case under winter rainfall patterns when rain water penetrates to deeper soil layers containing great amount of salts (McKenzie et al., 2004). Moreover, the removal of native vegetation from deep-rooted trees to shallow-rooted crops can led to a raise in the natural water table and consequently reduction in the remotion of salts from the surface. Similarly, these conditions can also occur with a return to higher rainfall afterwards long drought periods and subsequently improvement of the river flow, by which, also the salts accumulated on floodplains are distributed by the river (Blewett, 2012).

The previous cases exemplified some of the most common and difficult complications found in NSW soils, which at the same time become even more complex since the soil formation process here is very slow (Blewett, 2012; EPA, 2012). Clearly all these issues promote nutrient and physical limitations in these soils and therefore complications for agricultural practices. It has been reported in the NSW State of The Environment 2012 that from a state-wide basis view, 46% of NSW soils are in fair condition – with a 38% in good and 16% in poor conditions.

Consequently, there has been a moderate decline in the NSW soils ability to provide ecosystems services, including agricultural productivity (EPA, 2012).

Salinization, acidification and erosion together with a decline in the organic carbon and waterlogging have been reported as the most unsustainable managed soil degradation pressures in NSW. Soil degradation is considered one of the most difficult environmental management problem facing NSW these days, which principals causes rely on growing populations, increasing intensification of agriculture and degrading vegetation. Although it has been introduced some conservation practices and there are few rehabilitation initiatives ongoing (*e.g.* reduced tillage for soil erosion control) still a major effort is needed to put these particular issues under complete control (Blewett, 2012; EPA, 2012).

Land use

Undoubtedly, it is part of this research to consider differences in land use when analysing distribution of microbial diversity across NSW. This is why land use has been considered as an important factor when producing the sampling design of this investigation. Firstly, it is clearly necessary to differentiate either the microbial population living in lands managed differently to those living in natural or undisturbed ecosystems; additionally, this differentiation also allows us to elucidate the impacts of the land management over those soil microbial communities. Secondly, the vast majority of NSW land is used forestry, nature conservation. However, agriculture is one of the primary economic activities in NSW, and the land surface designated to

agricultural activities corresponds to an important portion of this terrain, as we can see on the basis of the land coverage in Figure 2-16.

In 2013-2014, the land managed by agricultural businesses comprised 58,3 million hectares (72,8% of NSW) distributed in approximately 44,000 farms. Figure 2-15 shows that agricultural lands are mainly used for grazing and cropping purposes whose areas scope respectively 78,7% and 16,8% of the total agricultural zone. Crops cultivation reached approximately 5,1 million hectares, recording in NSW the second largest area of lands prepared for crops in Australia (ABS, 2013).

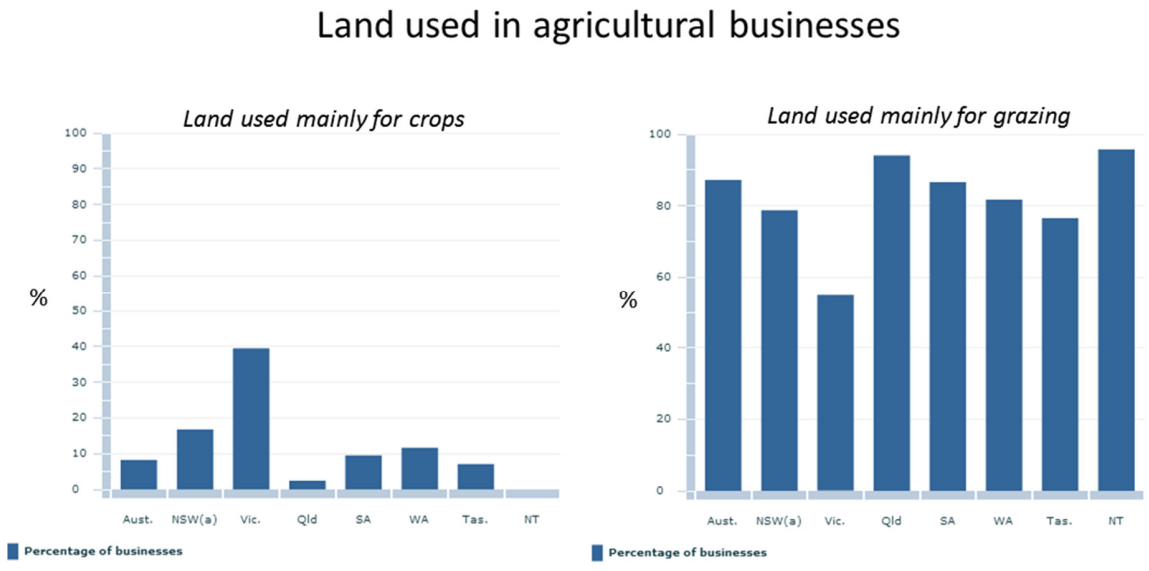


Figure 2−15. NSW agricultural land area (58, 3 million hectares) is mainly used by crops and grazing purposes.

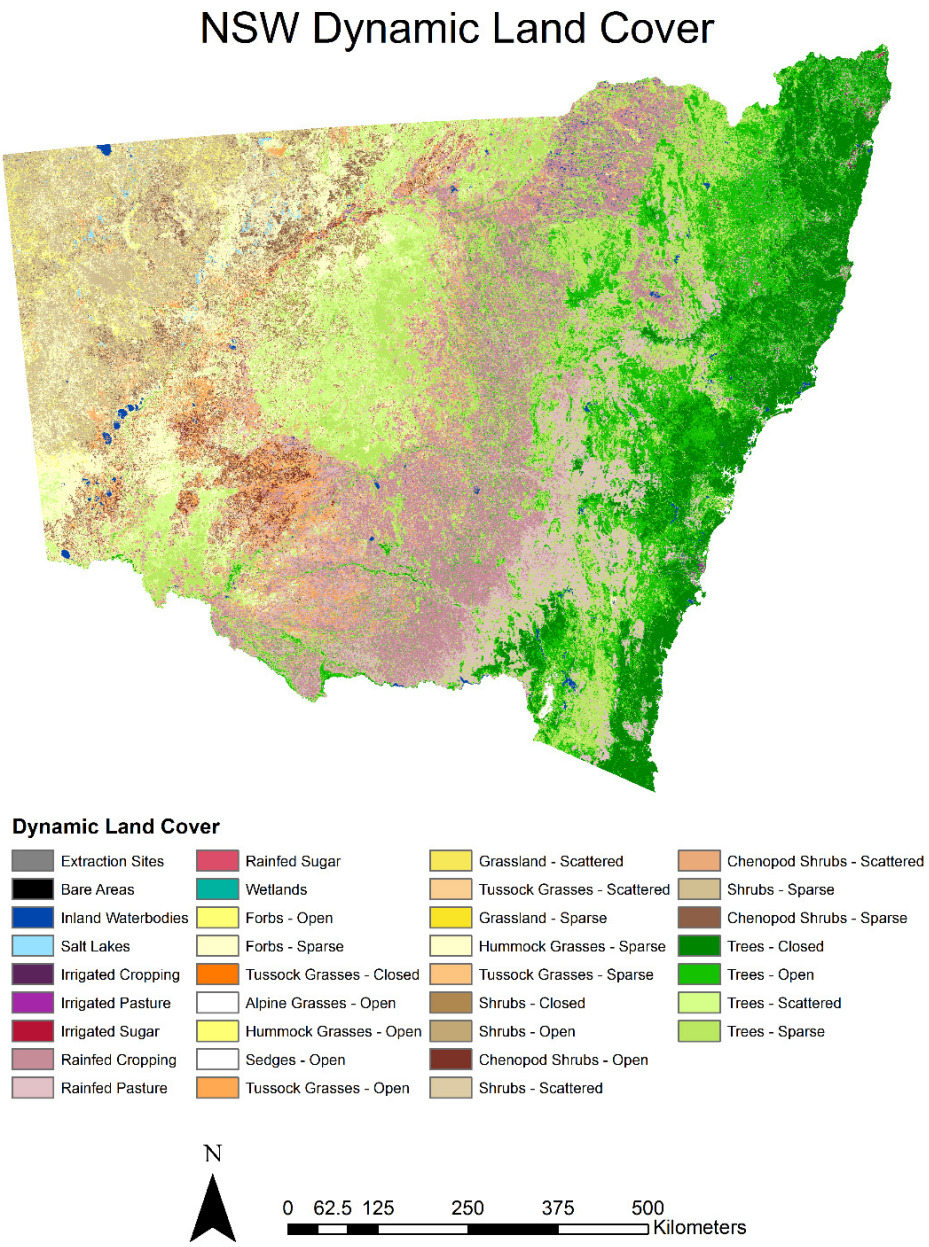


Figure 2–16. NSW DLCM according to the National Dynamic Land Cover Map for continental Australia[xix].

[xix] *The dynamic land-cover database is a time-series database based on an analysis of 16-day Enhanced Vegetation Index composites for the period 2000–08. It presents land-cover information based on the temporal behaviour of every 250x250-metre area of the country from April 2000 to April 2008. The classification scheme used to describe land-cover categories in the DLCM conforms to the 2007 ISO land-cover standard (19144-2).*

Grazing lands include natural pastures/grasslands, rangelands, woodland/shrub land, forested areas and swamps/wetlands and improved pastures. Cropping lands are primarily composed by winter crops such as wheat, barley and canola, which per year covers about of 5 million hectares and; summer crops such as cotton, rice maize, sorghum, soybeans and sunflowers, which each year can compromise nearly 700,000 hectares sowed. NSW also produce a range of horticultural crops including vegetables, fruit (pome, stone, berry and tropical), nuts, cut flowers and turf.

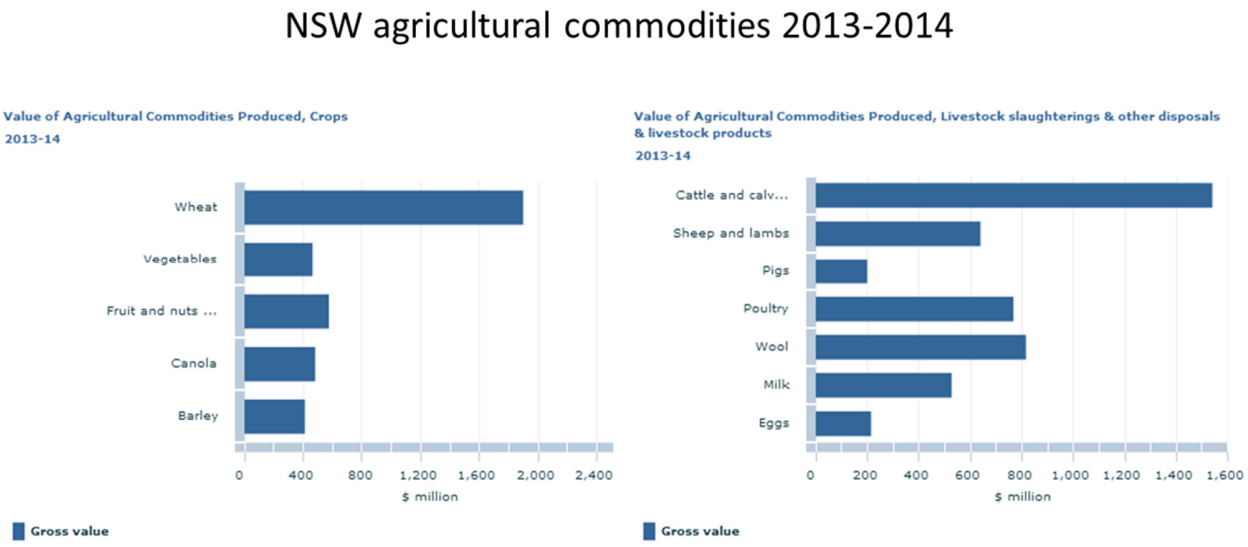


Figure 2−17. Commodities values of agricultural industry during period 2013−2014 in NSW (ABS, 2013).

On a commodity basis, wheat production is the most valuable contribution in the agricultural industry by representing nearly $1,900 million of the gross value of the agricultural commodities (Figure 2-17) (ABS, 2013). The major contribution to these statistics comes from the cropland area known as *Wheatbelt East Region,* one of the most important for the Australia grain

production. In this terms, it is noticeable that almost our complete $NS_{\text{-transect}}$ and some locations of the $WE_{\text{-transect}}$ encompass the area covered by this region.

Wheatbelt East Region corresponds to a relatively narrow band of land but very large in extension (17 million hectares). In this region takes place most of NSW grain agriculture but it is mostly occupies by white and hard wheat production (Figure 2-3). This terrain is largely characterized by fertile soils, temperate climate and sufficient rainfall gradient for rainfed production (ABARES, 2012; ABS, 2013). In this zone, rainfed agriculture prevail over irrigated areas and we can see in the dynamic land cover map the extended areas under rainfed cropping and pastures the Interior Lowland Division (GA, 2010).

Dynamic land cover map also exposes the diverse forms and distribution of NSW vegetation (Figure 2-16). There are a visible latitudinal pattern of the vegetation gradient, where western NSW is mostly characterized by open to sparse grasses and forbs; and conversely, the eastern NSW is significantly surrounded by closed to open trees, which is undoubtedly a response to the rainfall pattern. In general terms, NSW exhibits diverse types of vegetation varying from desert and rainforests to wetlands, heathland, grasslands, alpine herb fields, eucalypt forests and woodlands (GA, 2010).

On a state-wide basis, protected areas scope nearly 9,1% of NSW state, of which, national parks represent the major contribution with 6,4% followed by nature reserves with 1,19% (CAPAD, 2014). The proportion of agricultural lands aside for conservation or protection is 2,3%. Land

uses such as industrial, infrastructure, mining, and urban all together occupy no more than 2% of NSW (ABS, 2015b, 2013; EPA, 2012).

Nowadays, land management practices are broadly sustainable in NSW but the types of risks leading to soil degradation are variable across the state and there are some areas under high pressure caused by human-induced land uses practices. For instance, the irrigated areas of western slopes being degraded by salinity or those areas under sugar cane cropping on coastal acid sulphate soils (EPA, 2012).

Ecoregions, Bioregions and Biodiversity

The diversity of landscapes and climate found across NSW accommodate the presence of a wide variety of environments. An overall view of the main changes across these environments is better understood from an ecological perspective by integrating all the described environmental attributes (abiotic factors) shaping NSW together with the assemblage of living organisms and communities (biotic factors).

The most general view based on this ecological interaction is given by the terrestrial ecoregion classification. There are 14 ecoregion across the world and NSW encompasses six of the eight found in Australia. The NSW ecosystem diversity with respect to other Australia's states is clearly shown in the ecoregions map provided below (Figure 2-18). This ecoregions map also shows that the distribution of NSW ecoregions are very well defined along a latitudinal pattern. Western

NSW is characterized by Deserts and Xeric Shrublands in the north and Mediterranean Forests, Woodland and Shrub in the south. Instead, middle latitudes are mainly formed by Temperate Grasslands, Savannas and Shrublands. On its own, Eastern NSW is largely formed by Temperate Broadleaf and Mixed Forest having and considerable area of Tropical and subtropical Grasslands, Savannas and Shrublands sharing the border with Queensland's state in the north. Each of these ecoregion contains several biomes and these biomes may transcend ecoregion borders (DSEWPaC, 2012b). More specific features of these ecoregions are given in the table below:

A deeper view of those ecoregions is given by the Interim Biogeographic Regionalisation of Australia[xx] (DSEWPaC, 2012a). This is a system of regionalisation based on the dominant landscape attributes including climate, geomorphology, landform, lithology, and characteristics of flora and fauna. According to IBRA7 (2012), NSW contains 17 of the 89 biogeographic regions found all over Australia, from which approximately 8 bioregions are included in the target area of this study for the evaluation of soil microbial diversity (Figure 2-19).

On a biodiversity basis, it is noticeable that some significant bioregions are in our sites of sampling. For example, just surrounding the target area in northern NSW, are two locations identified on the list of Australia's 15 National Biodiversity Hotspots known as the Border Ranges and the Brigalow Belt (NARCliM, 2014).

[xx] *IBRA7 is the Australia National Reserve System's planning framework, the fundamental tool for identifying reservation targets and setting priorities to meet them. Vegetation community and land system mapping undertaken by the states and territories have been used to establish IBRA Region and Subregion Boundaries (DSEWPaC, 2012a).*

Table 5. Terrestrial ecoregions found across NSW (DSEWPaC, 2012b).

NSW ecoregion's	Description
Deserts and Xeric Shrublands	Evaporation exceeds rainfall in these ecoregions. Searing daytime heat gives way to cold nights because there is no insulation provided by humidity and cloud cover. Not surprisingly, the diversity of climatic conditions – though quite harsh – supports a rich array of habitats.
Mediterranean Forests, Woodlands and Shrubs	Only five regions in the world experience these conditions and whilst the habitat is globally rare, it features extraordinary biodiversity of uniquely adapted animal and plant species and the five areas collectively harbour well over 10 per cent of the Earth's plant species. Most plants are fire adapted, and dependent on this disturbance for their persistence.
Temperate Grasslands, Savannas and Shrublands	This ecoregion differs largely from tropical grasslands due to the cooler and wider annual temperatures as well as the types of species found here. Generally speaking, these regions are devoid of trees, except for riparian or gallery forests associated with streams and rivers. Positioned between temperate forests and the arid interior of Australia, the southeast Australian temperate savannas span a broad north–south swatch across Queensland, New South Wales and Victoria. Unfortunately, most of this ecoregion has been converted to sheep rearing and wheat cropping and only small fragments of the original eucalypt vegetation remains.
Temperate Broadleaf and Mixed Forests	Temperate forests experience a wide range of variability in temperature and precipitation. In regions where rainfall is broadly distributed throughout the year, deciduous trees mix with species of evergreens. Species such as Eucalyptus and Acacia typify the composition of the temperate broadleaf and mixed forests. In Australia, these temperate forests stretches from southeast Queensland to south Australia enjoying a moderate climate and high rainfall that give rise to unique eucalyptus forests and open woodlands. This biome has served as a refuge for numerous plant and animal species when drier conditions prevailed over most of the Australia continent. That has resulted in a remarkably diverse spectrum of organisms with high levels of regional and local endemism.
Tropical and Subtropical Grassland, Savannas and Shrublands	Large expanses of land in the tropics do not receive enough rainfall to support extensive tree cover. The tropical and subtropical grasslands, savannas, and shrublands are characterised by rainfall levels between 90–150 centimetres per year. While much of Australia is covered by grassland, savannah ecosystems are far more restricted – these ecosystems are limited to moister areas along the coast. Patches of dry rainforest with high species diversity also occur throughout the ecoregion.
Montane grasslands and Shrublands	This ecoregion includes high elevation (montane and alpine) grasslands and shrublands. In fact, this ecoregion in Australia is restricted to the montane regions of south–eastern Australia above 1300 metres. This region occupies less than three per cent of the Australian landmass and straddles the borders of the Australian Capital Territory, Victoria and New South Wales.

The first area is the most biologically diverse found in NSW, which is characterized by subtropical rainforest, wet sclerophyll forest, mountain headland and rocky outcrops. The second area is surrounded by large areas of eucalyptus and it was originally dominated by a vast community of Brigalow (*Acacia harpophylla*) (NARCliM, 2014).

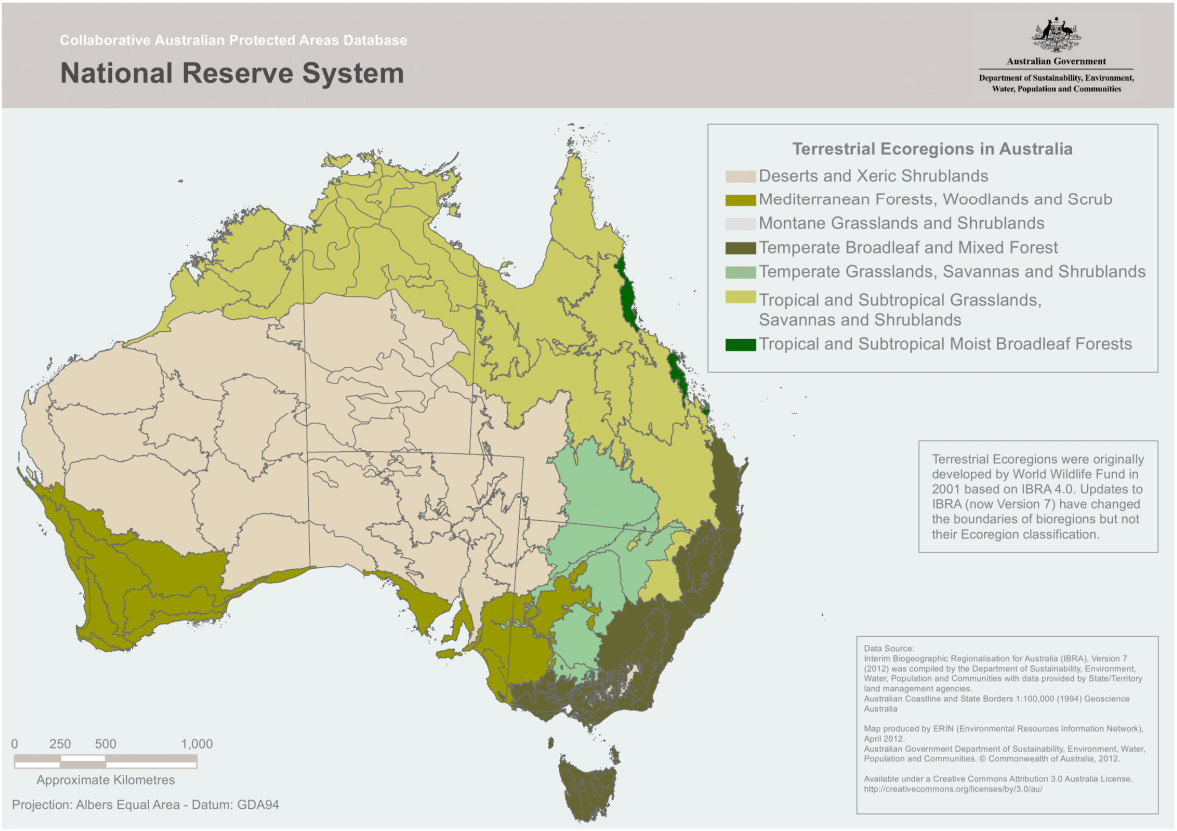


Figure 2-18. Australia's terrestrial ecoregions.

As it is seen in Figure 2-19, bioregions of western NSW ranges across sandy deserts (Simpson-Strzelecki Dunefields, Channel Country and Murray Darling Depression), riverine plains (Riverina and Darling Riverine Plains), rocky ranges (Mulga Lands, Broken Hill Complex

bioregions) and rolling downs (Cobar Peneplain). Towards the east of the State, there are lush rainforests (NSW North Coast, South East Corner), rugged mountains (Sydney Basin, New England Tableland, Australian Alps, and South Eastern Highlands), undulating ranges (Brigalow Belt South, Nandewar) and fragile wooded grasslands (NSW South Western Slopes). Specific details of each these bioregions can be found in the Office of Environment & Heritage of the NSW Government (DSEWPaC, 2012a).

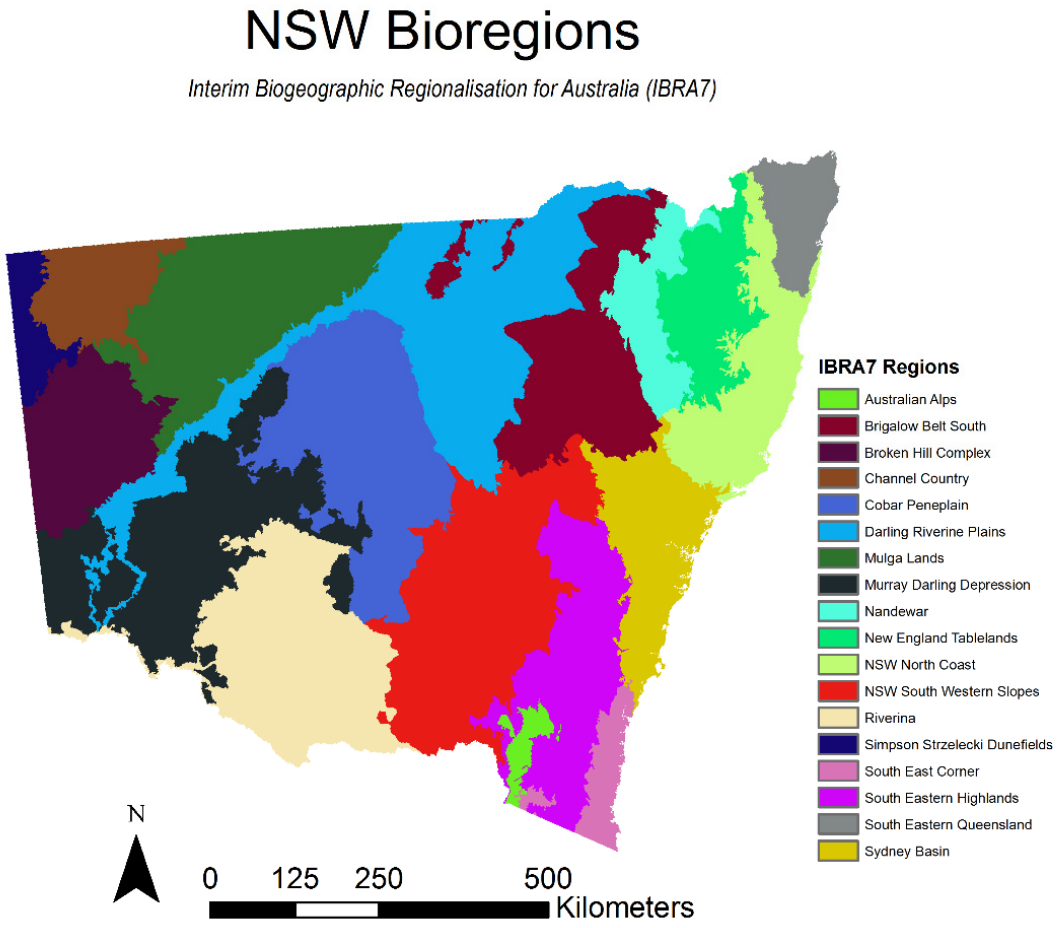


Figure 2−19. NSW bioregions according to the Interim Biogeographic Regionalisation for Australia (DSEWPaC, 2012a)

DESIGNING THE TRANSECTS

A geographic information system (GIS) using ArcGIS 10.0 software (ESRI, 2011) was developed to design *in silico* the sampling area in NSW. It has been designed north-south ($NS_{\text{-transect}}$) and west-east ($WE_{\text{-transect}}$) transects to represent their respective latitudinal and longitudinal agroecological gradients in this area. Representative sampling sites were allocated along both transects, defined based on a set of climate, soil and land cover references across NSW. Two different land-use ecosystems both natural ecosystems and cropping ecosystems were differentiated per each sampling site based on dynamic land cover information. All the environmental references discussed above were extracted from raster graphics images and spatial information datasets which were pre-processed to be carried out at WGS_1984_UTM_Zone 56S spatial reference system. In general terms, the entire design procedure for both transects was complete in three phases, the details are described below.

NS-transect design

$NS_{\text{-transect}}$ extend 900 km from Queensland border in the north to the Victorian border in the south. This transect was extracted as an equivalent contour line of the 550 mm average annual rainfall isohyet to represent a longitudinal agroecological gradient but also to encompass an important rainfed agricultural zone (Figure 2-20). The 550 mm isohyet and respective contour line were processed in ArcGIS from the raster image of the NSW average annual rainfall at ~ 2.5

km resolution provided by Australian Bureau Meteorology (BOM, 2009). As a result, owing to the longitudinal position of the isohyet, $NS_{\text{-transect}}$ runs through the transitional zone between Interior Lowland and *Eastern Uplands* - described and referred above as agricultural plains and *Wheatbelt* region.

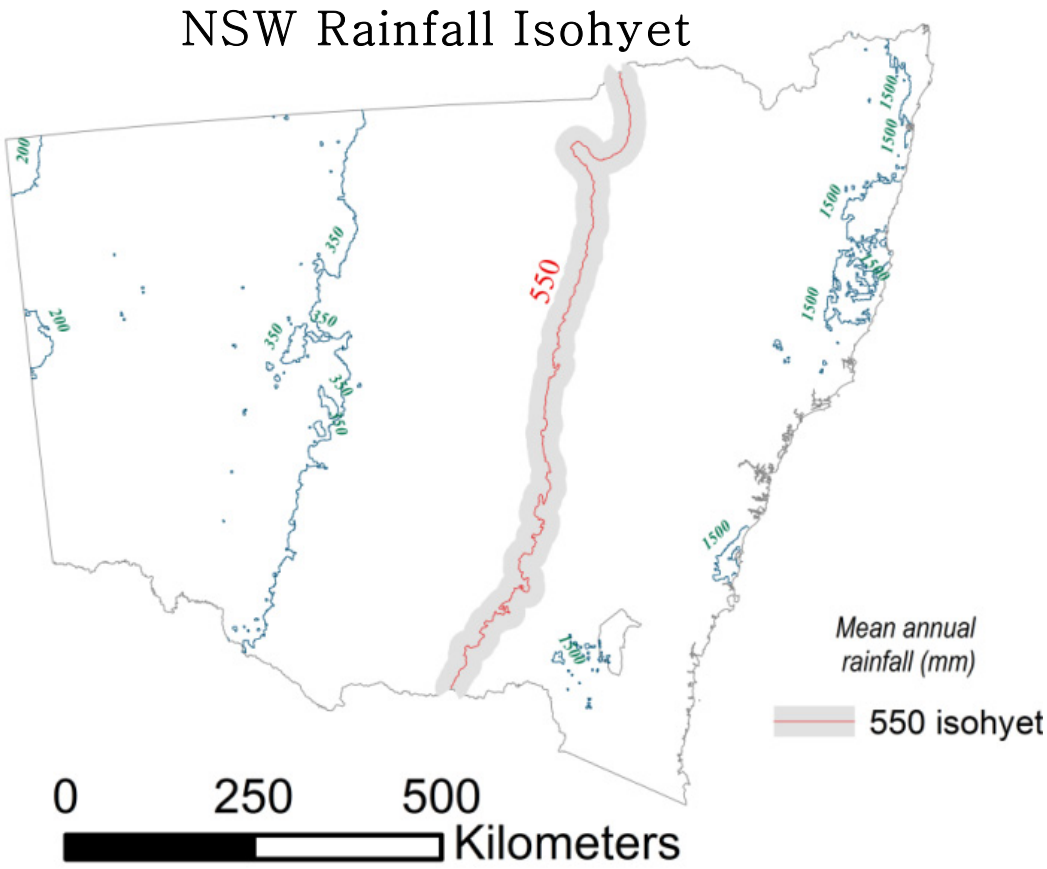


Figure 2–20. Contour lines of NSW rainfall isohyets.

WE-transect design

On the other hand, the $WE_{-transect}$ extend 930 km from the coastal area of Coffs Harbour (about 220 km south of the Queensland border) and runs west to the isolated settlement of Wanaaring located in outback NSW. This transect was designed to encompass a large latitudinal environmental gradient including a maximum threshold of the rainfall gradient. For this environmental variability it was very important when designing the $WE_{-transect}$ to consider the access to sampling points (*e.g.* highways and roads) in the far western areas of NSW (the outback) (Figure 2-21). As a result, summarizing all of the previous aspects, $WE_{-transect}$ was positioned almost perpendicular to the $NS_{-transect}$.

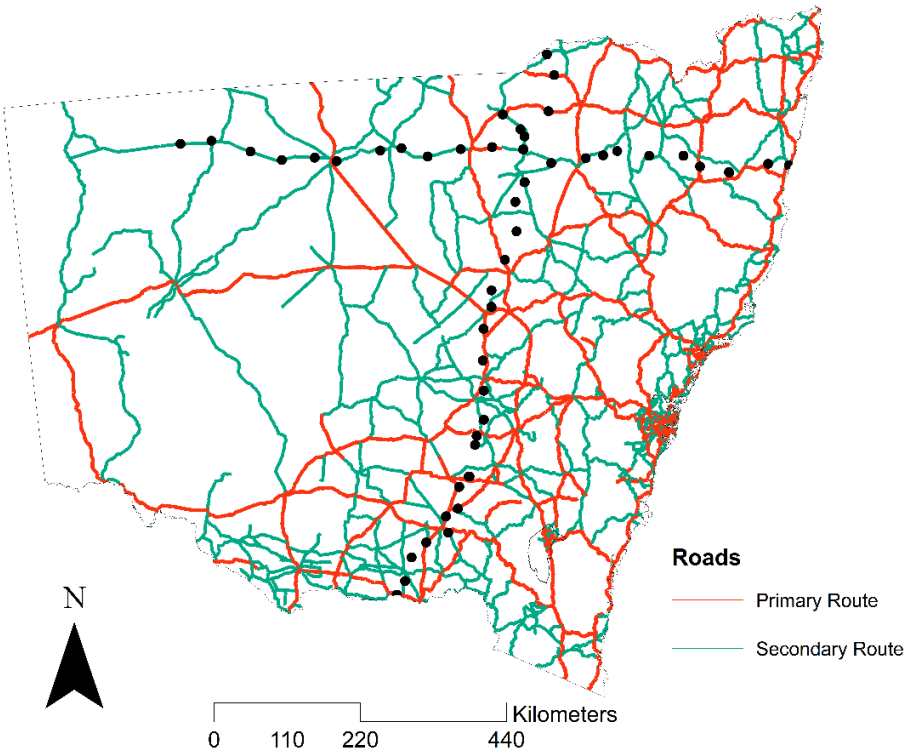


Figure 2-21. Principal highways allowing the sampling the western part of $WE_{-transect}$.

Selecting the sampling sites

Sampling sites were allocated at a separation distance of ~ 50 km along the previously traced contour lines *i.e.*, $NS_{-transect}$ and $WE_{-transect}$. These sampling sites were carefully selected from the most representative environmental zones within a 20 km radius at each sampling location.

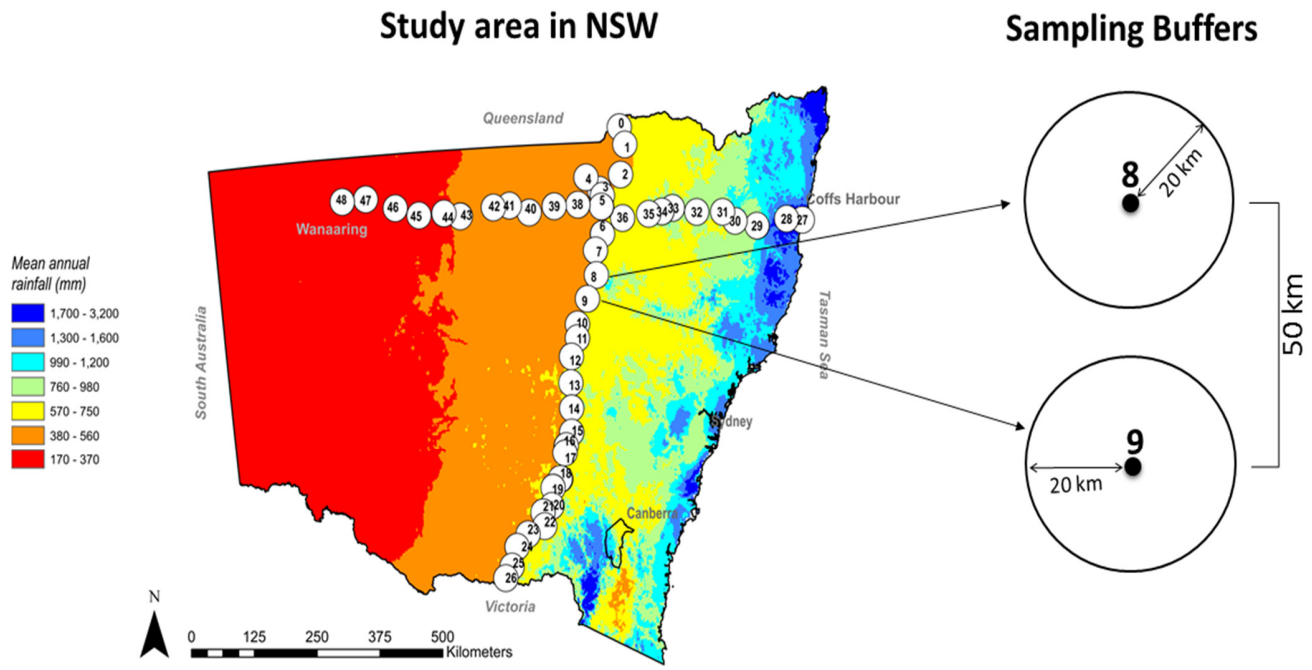


Figure 2–22. Sampling buffers distributed along $NS_{-transect}$ and $WE_{-transect}$.

This selection process started by plotting sampling buffers of 20 km radius which centroids were placed at a separation distance of 50 km to each other along the representative lines of the transects. In total, 27 sampling buffers were distributed along $NS_{-transect}$ and 22 along $WE_{-transect}$, completing 49 sampling locations composing the entire study area (Figure 2-22). Each of those

buffers were individually analysed in term of their environmental variability to select representative zones. These zones were defined based on a frequency distribution analysis of key environmental variables such as climate, mineralogy, land cover, soil pH, salinity, soil type, among others. This information was processed as input Arc/Info grid data for geostatistical analysis obtained from the Geophysical Archive Data Delivery System (GADDS), the Geoscience Australia (GA) and Australian Bureau Meteorology (BOM, 2009).

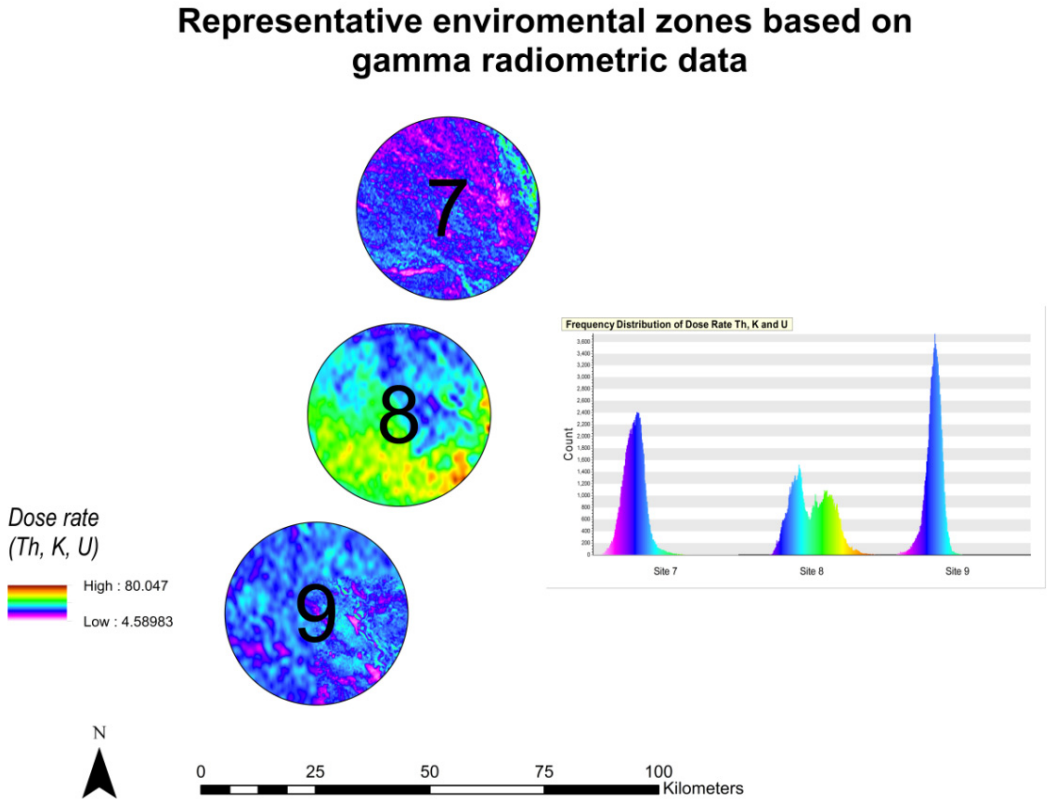


Figure 2−23. Representative zones selected by frequency distribution analysis. This example shows the selection procedure based on the gamma radiometric data in the sites 7, 8 and 9 of the $NS_{-transect}$.

Once defined the representative environmental zones, the final sampling sites were randomly selected from areas by selecting contiguous natural or undisturbed ecosystems (*e.g.* forests, woodlands or natural grasslands) and disturbed or managed ecosystems (*e.g.* cropping, grazing lands). To ensure the same soil composition under both land uses (*e.g.* natural and cropping ecosystems), the selection process was performed using both the Dynamic Land Cover dataset of Australia at 250 m resolution and The Radiometric Map of Australia dataset at 100 m resolution (GA, 2010; GADDS, 2010). Finally, the soil samples for microbial and physicochemical analysis in this investigation have been taken from natural and cropping ecosystems at each of the 49 sampling sites. Details about the sampling campaigns and collection protocol are explained below.

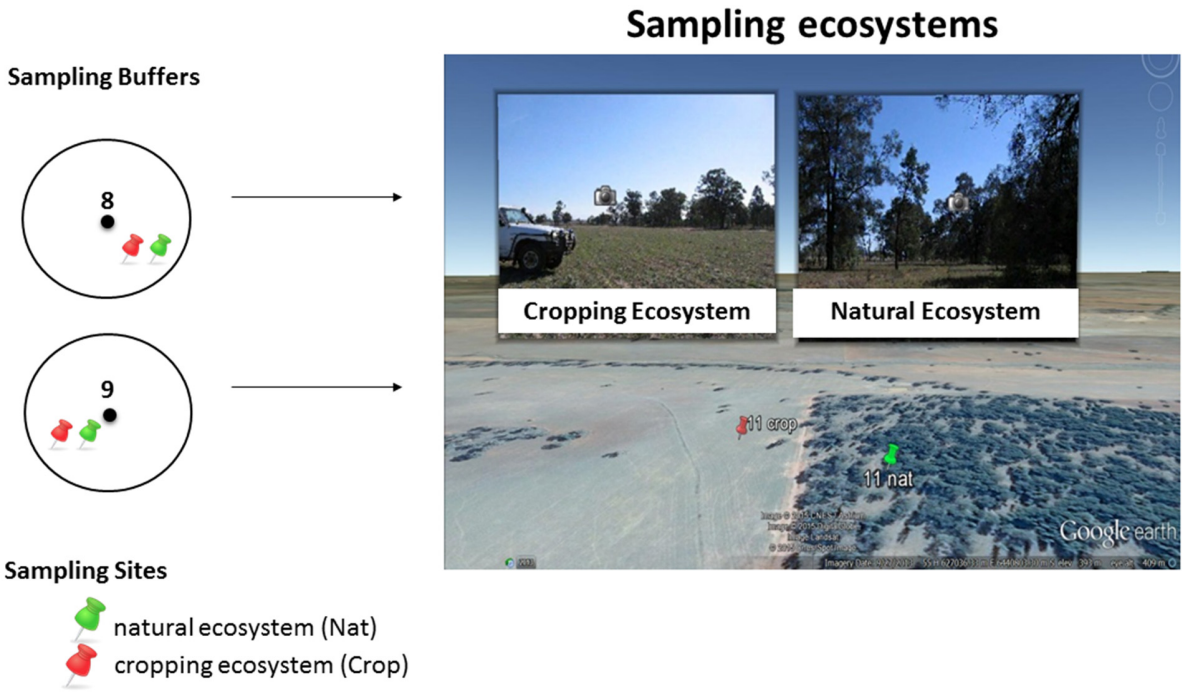


Figure 2–24. Two sampling ecosystems representing both: natural (undisturbed) and cropping (disturbed) land uses.

FIELDTRIPS AND SOIL SAMPLES COLLECTION

SAMPLING CAMPAIGNS

The complete study area was sampled in seven fieldtrip campaigns. Sampling times were selected to ensure sampling at season's time when microbial population growth was expected to be high. Thus, the sampling campaigns were carried out during early autumn in 2013 along $NS_{-transect}$ and early autumn in 2014 along $WE_{-transect}$ (Table 2.6). The climate conditions during the sampling campaigns are summarized in term of the observed total rainfalls (Figure 2-25) and extremes maximum (Figure 2-29) and minimum (Figure 2-30) temperatures during the sampling seasons in 2013/2014.

Table 2.6. Soil samples collection campaigns.

Transect	Sampling sites	Campaigns	Date	Season	Samples
$NS_{-transect}$	0–26	5	Mar–Apr 2013	Autumn	426
$WE_{-transect}$	27–48	2	Mar–Apr 2014	Autumn	264

Soil samples were collected for both microbial and physicochemical analyses following in each case the protocols specified below along all the sampling sites throughout $NS_{-transect}$ and $WE_{-transect}$.

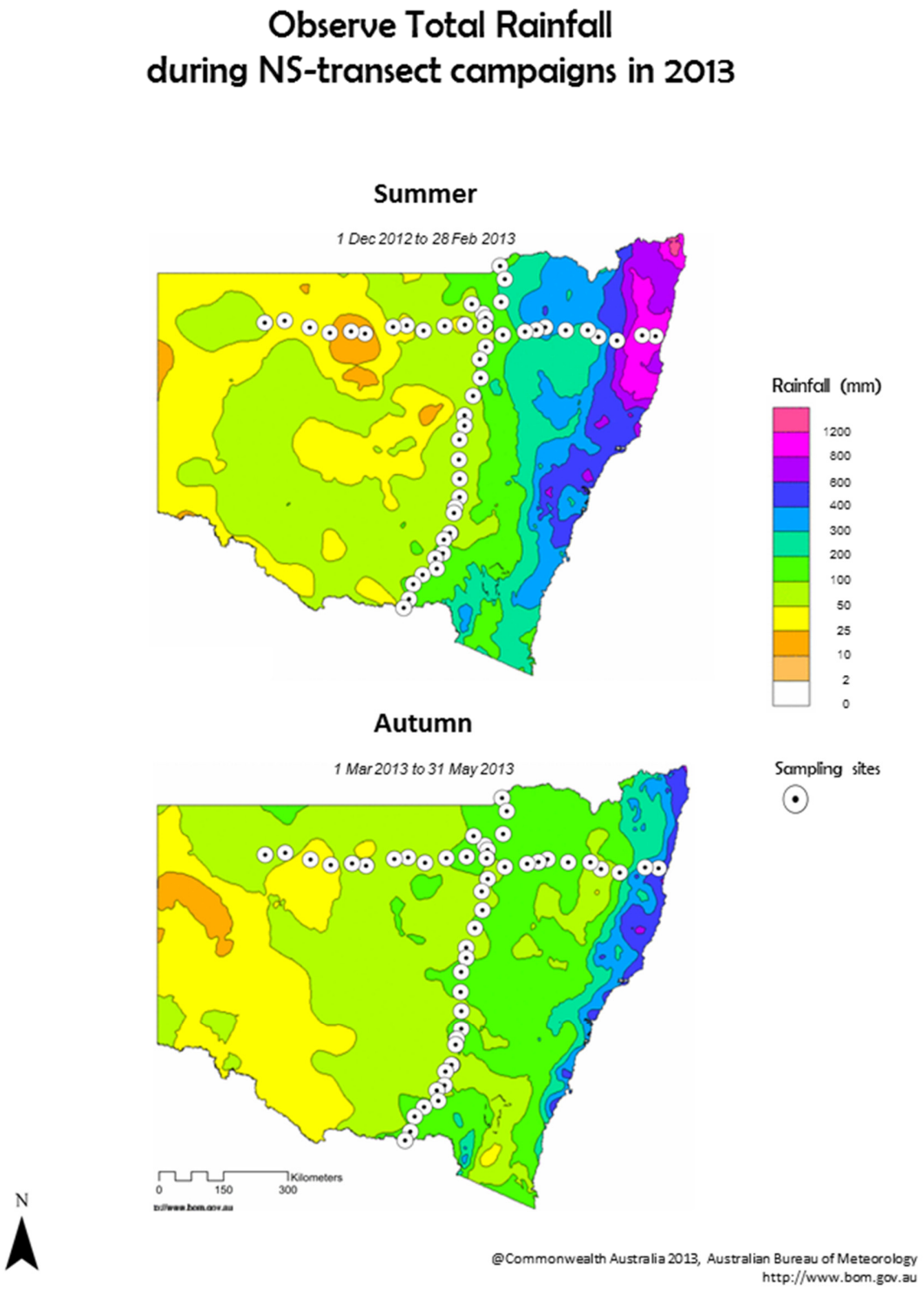


Figure 2-25. Total rainfall before and during sampling campaigns along NS-transect in 2013.

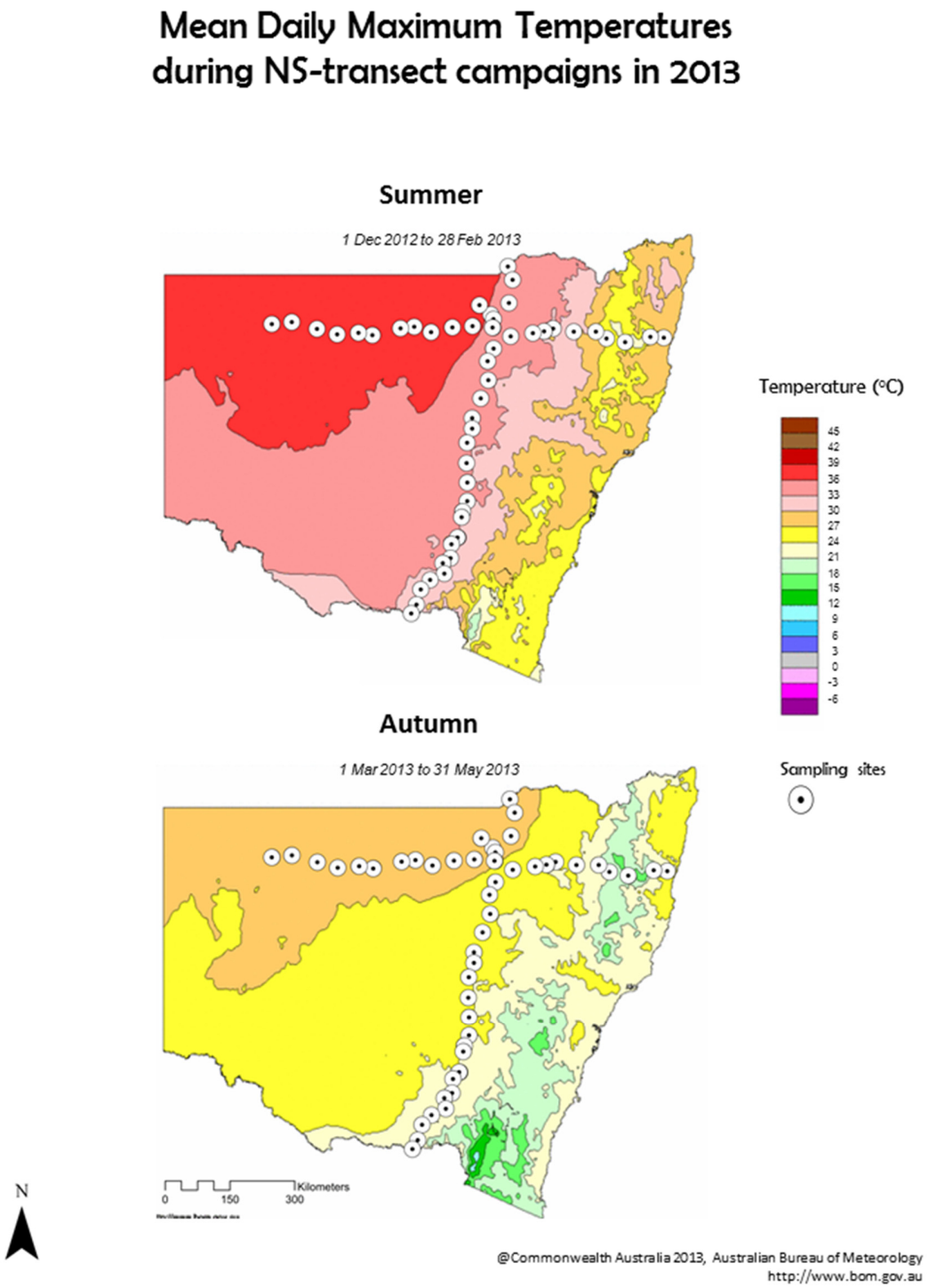


Figure 2-26. Maximum temperatures before and during sampling campaigns along $NS_{-transect}$ in 2013.

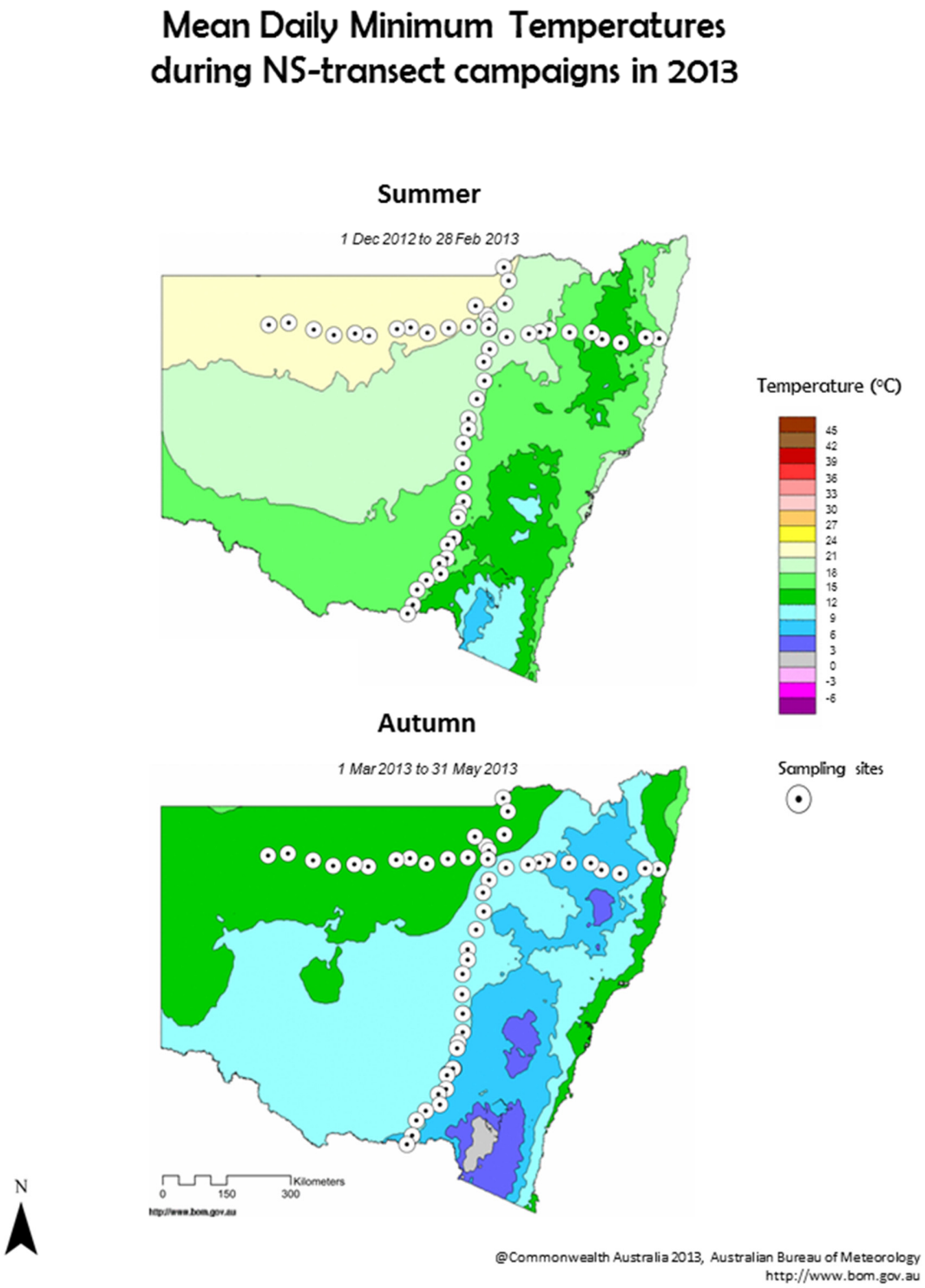


Figure 2-27. Minimum temperatures before and during sampling campaigns along $NS_{-transect}$ in 2013.

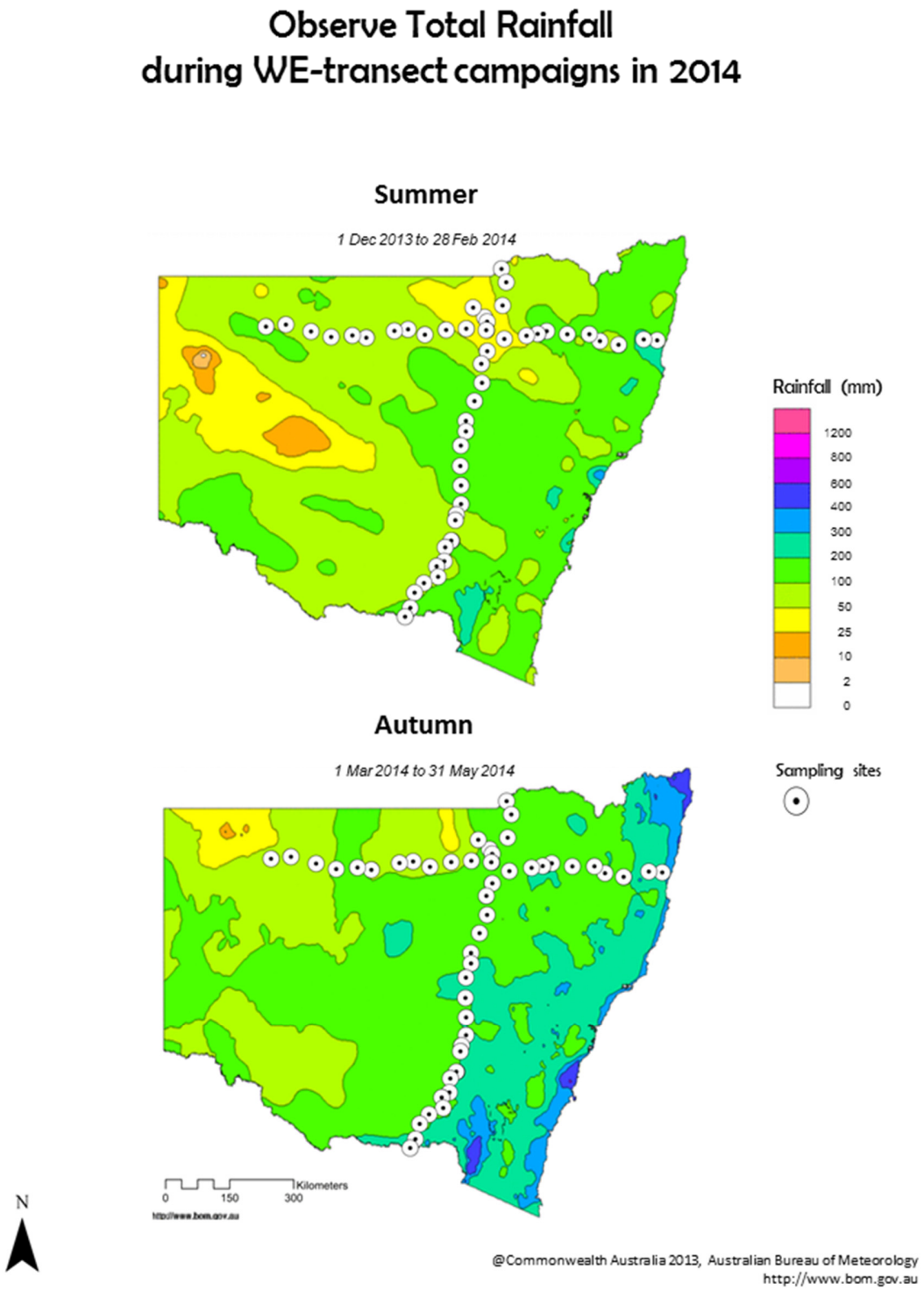


Figure 2-28. Total rainfall before and during sampling campaigns along WE-transect in 2014.

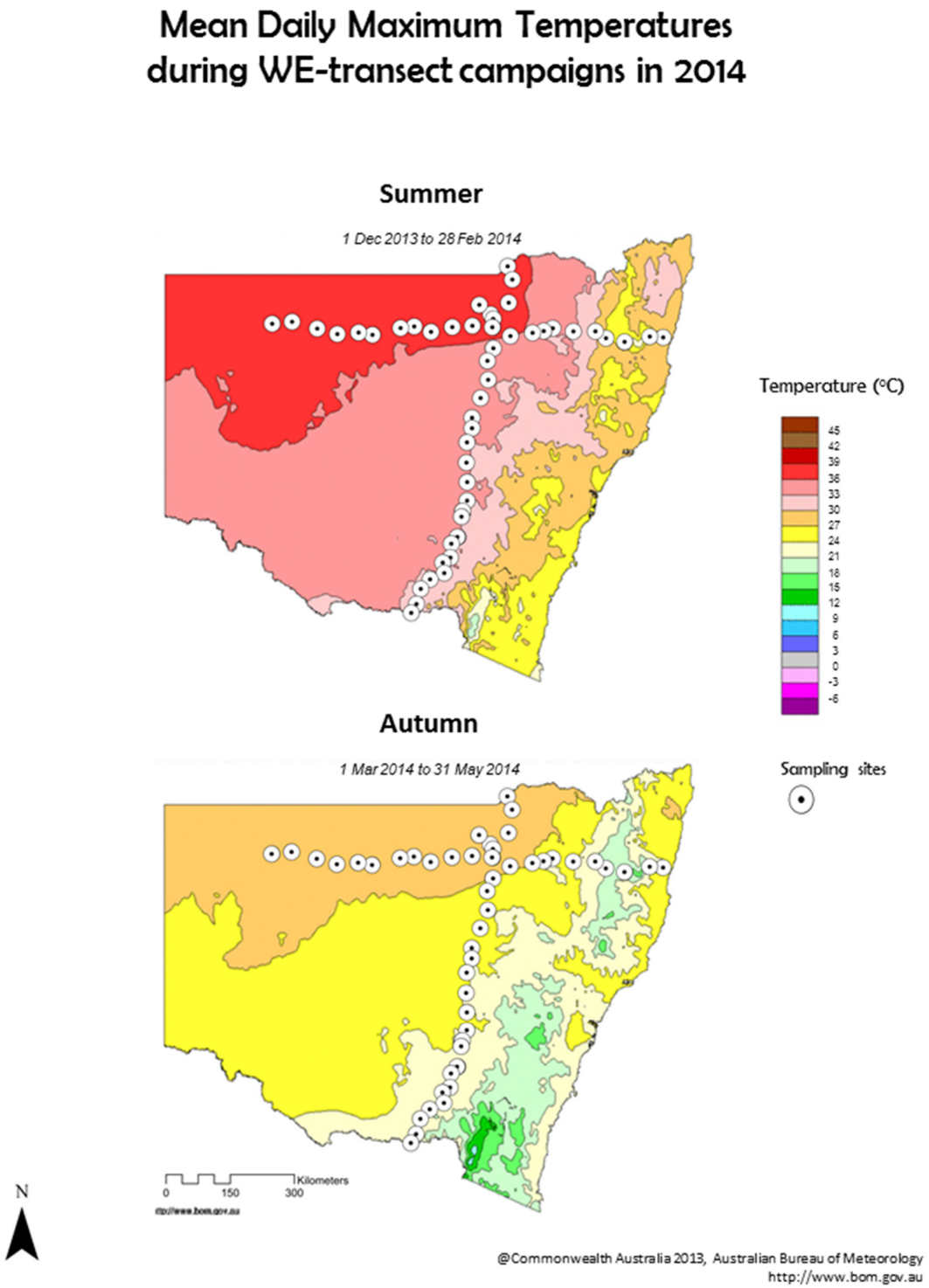


Figure 2-29. Maximum temperatures before and during sampling campaigns along WE-transect in 2014.

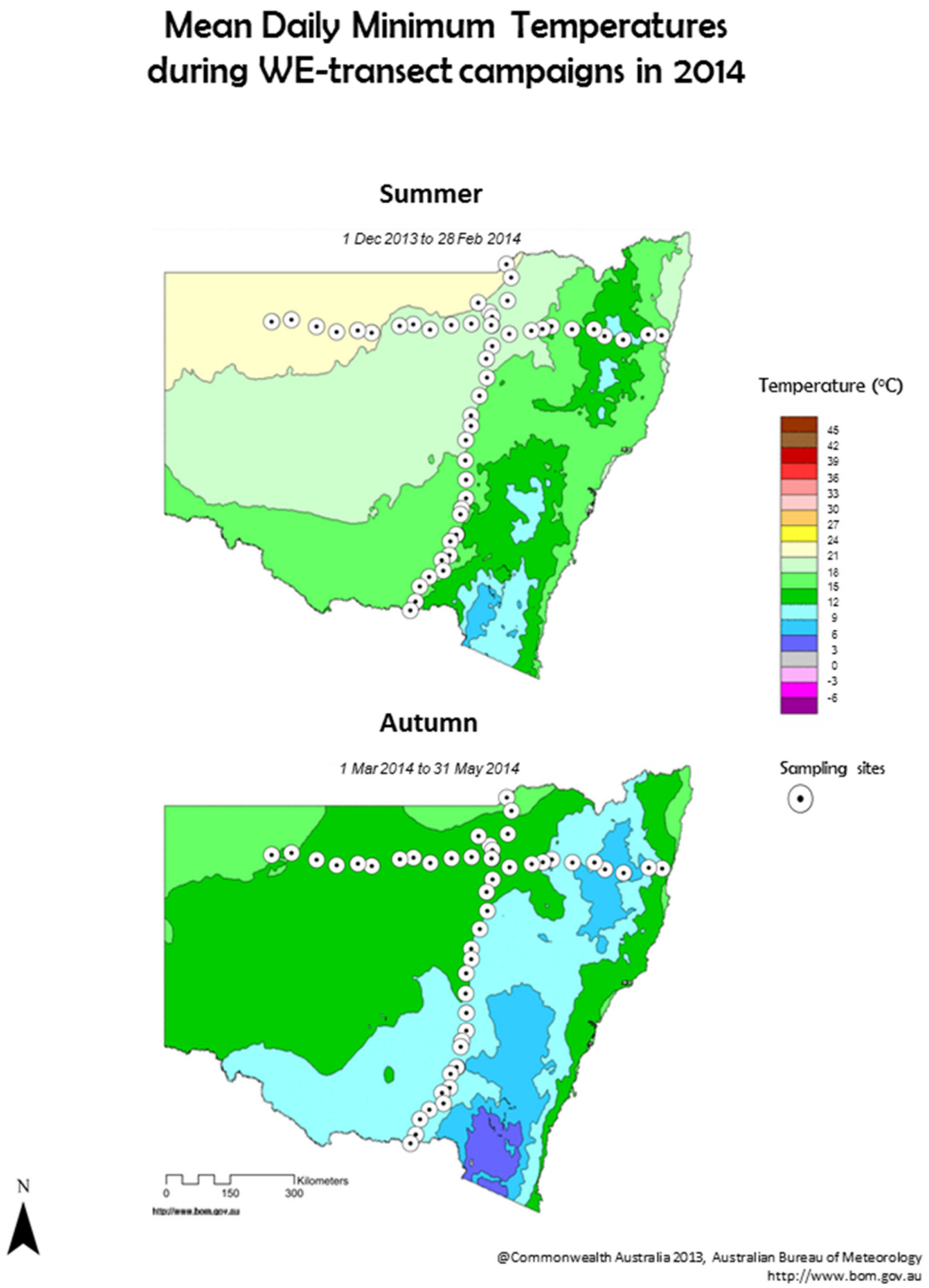


Figure 2-30. Minimum temperatures before and during sampling campaigns along WE-transect in 2014.

SAMPLE COLLECTION PROTOCOL

As explained above, the soil samples for microbial and physicochemical analysis have been taken from both natural/undisturbed and disturbed/managed ecosystems beside the 49 sampling sites composing the entire study area. In general terms, most of the soil samples obtained in natural ecosystems come from woodlands, forest (*e.g.* State Forests and National Parks) or natural grasslands. In the case of the soil samples obtained in managed ecosystems, most of them come from cropping ecosystems - primarily cereals (*e.g.* oats, wheat) - or natural/improved pastures for grazing purposes. However, it has to be highlighted that due to the lack of managed land at the extreme west of the $WE_{\text{-transect}}$, the protocol applied in these sampling sites (#46-#48) was based on the presence/absence of native vegetation coverage. In these cases, most of the natural ecosystems correspond to sparse shrublands areas while most of their converse conterminous sampled were bare soils ecosystems under serious salinization pressure. Below are shown few examples displaying the differences between the conterminous land uses and management within sampling sites (Figure 2-33 to Figure 2-35).

Each ecosystem per sampling site was sampled within a 1 meter square at both 0-5 and 5-10 cm depth. Soil sampled for microbial analyses was taken in three replicates equally distributed into the referred sampling square at each depth (Figure 2-31). For each replicate, it was extracted approximately 50 gr of soil into sterile falcon tubes (50 ml). These falcon tubes were immediately stored into refrigeration at -4 °C and kept in such conditions until the end of the respective sampling campaign for about four to five days. Finally, these samples were stored at -21 °C into

the freezer installations belonging to the Soil Security Laboratory of The University of Sydney.

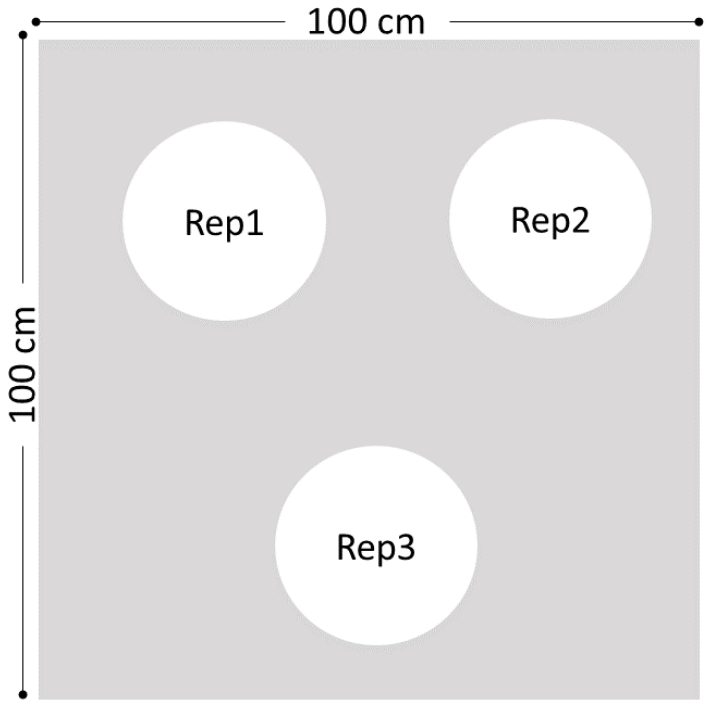


Figure 2−31. Three replicates were sampled within 1 meter square from each ecosystems per sampling site.

In the case of soil samples for physicochemical analyses, two replicates of 500 g of soil were taken from each depth. These replicates were composited by collecting equally distributed soil subsamples from the interior of the square meter. In addition to these replicates, a cylinder of 250 cc for other soil physical evaluations such as moisture, bulk density and porosity. To also validate soil parity between natural and cropping ecosystems, three replicates of soil profiles were extracted for each of these ecosystems in each site, using cores of 5 cm diameter and 1 m depth (Figure 2-32). These soil samples have been kept at 4 °C.

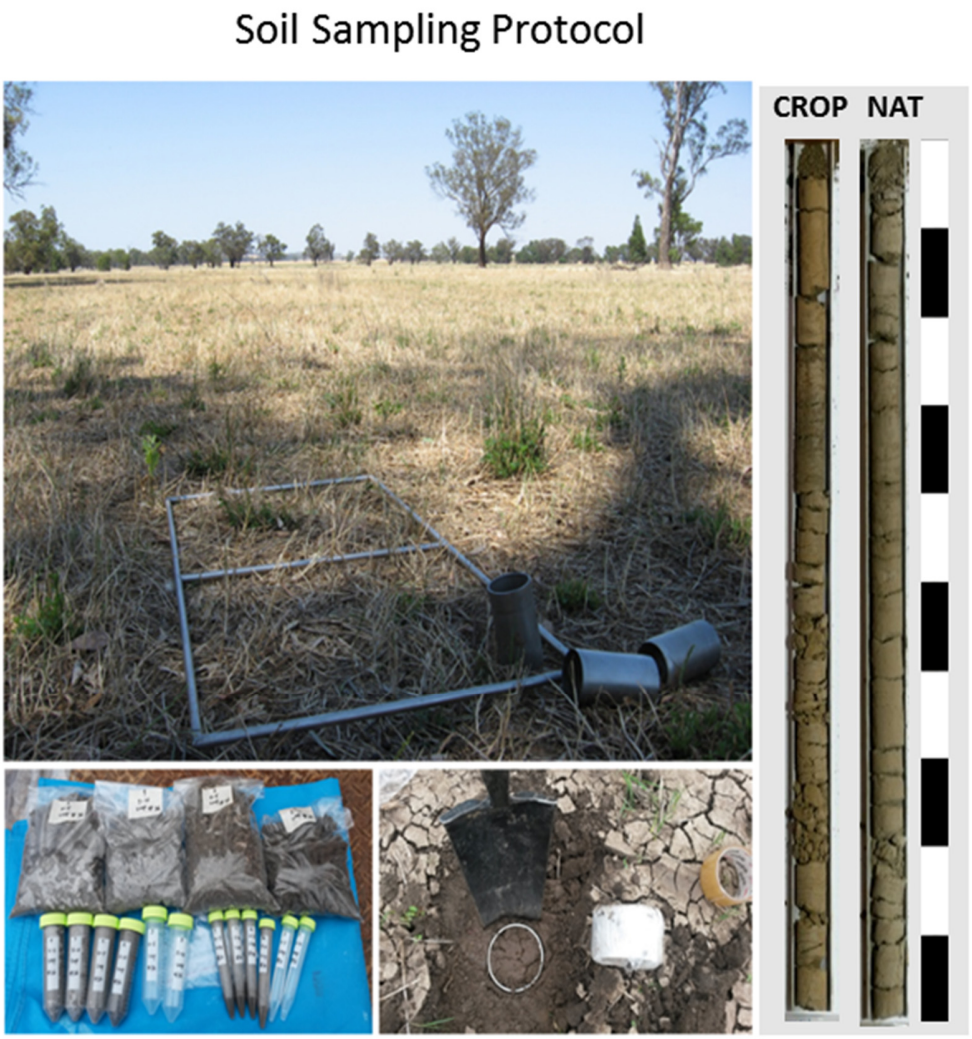


Figure 2-32. Soil samples for physicochemical and microbial analyses.

All the sampled sites, their specific locations and the specific ecosystems sampled are summarized in Table 2.7 and Table 2.8 for $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$, respectively.

Table 2.7. Sampling sites locations along the NS–transect.

# Site	Ecosystem	Latitude	Longitude	Elevation (m)	Suburb	Postcode
0	NAT	-28.7718063	149.3585040	165.4	Mungindi	2406
	CROP	-28.7721425	149.3583682	166.5		
1	NAT	-29.0551005	149.4612751	169.0	Garah	2405
	CROP	-29.0549532	149.4614359	170.9		
2	NAT	-29.5412864	149.3564663	165.4	Moree	2400
	CROP	-29.5414905	149.3566568	168.0		
3	NAT	-29.7740409	148.9116398	153.1	Rowena	2387
	CROP	-29.7739197	148.9112859	153.1		
4	NAT	-29.5664966	148.6448578	144.2	Collarenebri	2833
	CROP	-29.5647977	148.6465126	144.7		
5	NAT	-29.8725106	148.9704932	158.8	Rowena	2387
	CROP	-29.8716502	148.9707143	154.0		
6	NAT	-30.4926763	148.9443361	190.0	Gwabegar	2356
	CROP	-30.4919885	148.9487081	189.5		
7	NAT	-30.7548639	148.7919094	222.6	Urawilkie	2829
	CROP	-30.7526288	148.7917426	218.8		
8	NAT	-31.1540370	148.7923188	304.0	Coonamble	2829
	CROP	-31.1529320	148.7919938	301.4		
9	NAT	-31.5311823	148.5905584	262.8	Curban	2827
	PAST	-31.5306516	148.5889802	262.0		
10	NAT	-31.9387398	148.3594292	246.4	Kickabil	2830
	PAST	-31.9389310	148.3584164	246.4		
11	NAT	-32.1608406	148.3480938	266.2	Burroway	2821
	CROP	-32.1596554	148.3467296	262.2		
12	NAT	-32.4522649	148.2042657	268.9	Narromine	2821
	CROP	-32.4534246	148.2056110	261.1		
13	NAT	-32.8808143	148.1701455	294.6	Peak Hill	2869
	CROP	-32.8803075	148.1703313	295.8		
14	NAT	-33.2883890	148.1660953	285.5	Daroobalgie	2870
	CROP	-33.2904854	148.1656721	285.5		
15	NAT	-33.6807256	148.1444718	270.3	Glenelg	2810
	CROP	-33.6810327	148.1424694	268.2		
16	NAT	-33.8953773	148.0124404	375.7	Caragabal	2810
	CROP	-33.8962314	148.0160518	374.9		
17	NAT	-34.0130477	147.9827215	285.7	Bimbi	2810
	CROP	-34.0144474	147.9802251	280.6		
18	NAT	-34.4387885	147.8637494	281.0	Stockinbingal	2725
	CROP	-34.4368808	147.8515938	283.9		
19	NAT	-34.5698339	147.6896520	347.2	Combaning	2666
	CROP	-34.5722478	147.6885708	342.7		
20	NAT	-34.8623514	147.6463222	339.5	Wantiool	2663
	CROP	-34.8624577	147.6466840	343.2		
21	NAT	-34.9577616	147.4464569	224.0	Yathella	2650
	PAST	-34.9601610	147.4437973	220.0		
22	NAT	-35.1814365	147.4674019	246.0	Forest Hill	2651
	CROP	-35.1814085	147.4676434	245.0		
23	NAT	-35.2959704	147.0950706	265.6	The Rock	2655
	PAST	-35.2959617	147.0946527	260.1		
24	NAT	-35.4849928	146.8389483	250.1	Munyabla	2658
	CROP	-35.4849682	146.8407894	251.6		
25	NAT	-35.7986983	146.7074469	230.9	Brocklesby	2642
	CROP	-35.7968831	146.7062362	218.0		
26	NAT	-35.9754185	146.5562907	145.3	Howlong	2643

NAT: natural vegetated ecosystem (*e.g.* woodland, forest, grassland)
CROP: cropping ecosystem (*e.g.* cereals)
PAST: natural OR improved pastures ecosystem (*e.g.* intense grazing)
BARE: bare soil

Table 2.8 Sampling sites along the WE–transect.

# Site	Ecosystem	Latitude	Longitude	Elevation (m)	Suburb	Postcode
27	NAT	-30.32509664	153.0897726	34.2	Coffs Harbour	2450
	CROP	-30.3252499	153.0900016	43.1		
28	NAT	-30.30699219	152.7637639	795.3	Megan	2453
	PAST	-30.3071181	152.7635244	803.8		
29	NAT	-30.4202587	152.1521279	1115.6	Wollomombi	2350
	PAST	-30.41674504	152.1488162	1093.0		
30	NAT	-30.33575984	151.6975561	1351.2	Black Mountain	2365
	PAST	-30.33544418	151.6966657	1345.8		
31	NAT	-30.18700554	151.4404438	1167.8	Brushy Creek	2365
	PAST	-30.18736662	151.4412171	1161.4		
32	NAT	-30.18098279	150.9063981	867.8	Bundarra	2359
	PAST	-30.18113583	150.9058134	866.9		
33	NAT	-30.107111	150.4081673	387.9	Upper Horton	2347
	PAST	-30.10719358	150.4077813	390.5		
34	NAT	-30.1620034	150.1883063	515.7	Rocky Creek	2390
	PAST	-30.1599874	150.1855613	496.9		
35	NAT	-30.1923547	149.9139568	304.1	Narrabri	2390
	CROP	-30.19218969	149.9136716	301.5		
36	NAT	-30.24651213	149.3739561	189.8	Wee Waa	2388
	CROP	-30.24657154	149.3726693	188.1		
37	NAT	-30.05049875	148.9485179	161.6	Burren Junction	2386
	CROP	-30.04794325	148.9419656	166.0		
38	NAT	-30.00164127	148.4628718	144.2	Walgett	2832
	CROP	-30.00146054	148.4596451	142.2		
39	NAT	-30.0100543	147.9705799	129.0	Walgett	2832
	CROP	-30.01038494	147.9722421	128.5		
40	NAT	-30.08881491	147.4520964	128.9	Walgett	2832
	CROP	-30.09018491	147.4525904	127.0		
41	NAT	-29.96011591	147.0514788	124.2	Brewarrina	2839
	CROP	-29.9595831	147.0523699	118.8		
42	NAT	-29.97592125	146.7209114	118.0	Brewarrina	2839
	CROP	-29.9765275	146.7216664	117.3		
43	NAT	-30.08306944	146.0359298	109.0	Bourke	2840
	CROP	-30.08584169	146.0337342	110.9		
44	NAT	-30.01931213	145.6998976	111.2	Bourke	2840
	PAST	-30.01892609	145.7019245	109.2		
45	NAT	-30.02173797	145.1885574	122.7	Gumbalie	2840
	PAST	-30.02163929	145.1892538	112.3		
46	NAT	-29.87695365	144.7153935	102.0	Wanaaring	2840
	BARE	-29.87688318	144.7149707	105.0		
47	NAT	-29.69403189	144.1244693	106.5	Wanaaring	2840
	BARE	-29.69500263	144.1247309	111.7		
48	NAT	-29.70352057	143.6481657	166.9	Wanaaring	2840
	BARE	-29.70346023	143.6448682	163.8		

NAT: natural vegetated ecosystem (*e.g.* woodland, forest, grassland)
CROP: cropping ecosystem (*e.g.* cereals)
PAST: natural OR improved pastures ecosystem (*e.g.* intense grazing)
BARE: bare soil

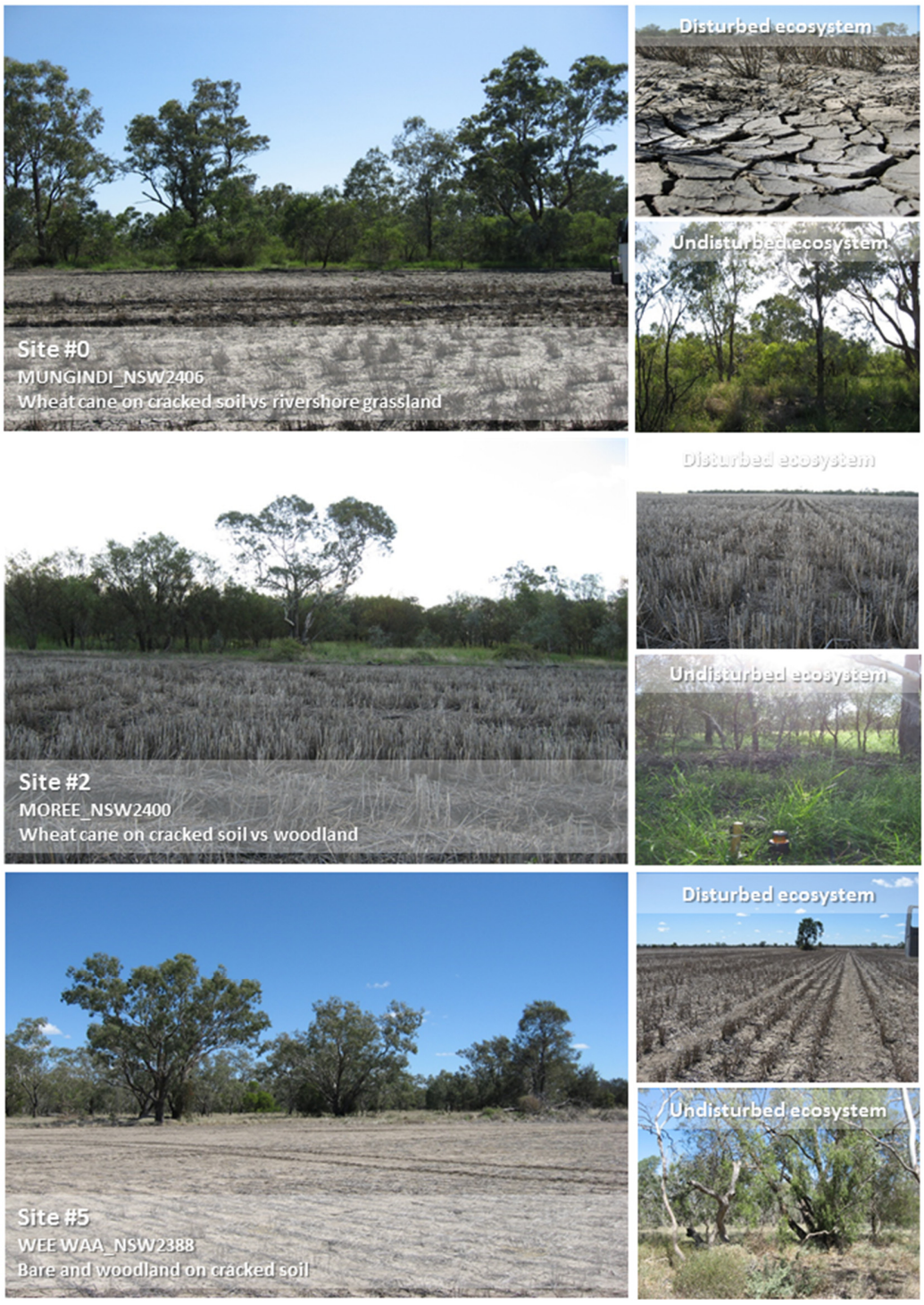


Figure 2-33. Some of the sampling sites sampled showing differences between undisturbed and disturbed ecosystems sampled in the north of NS-transect.

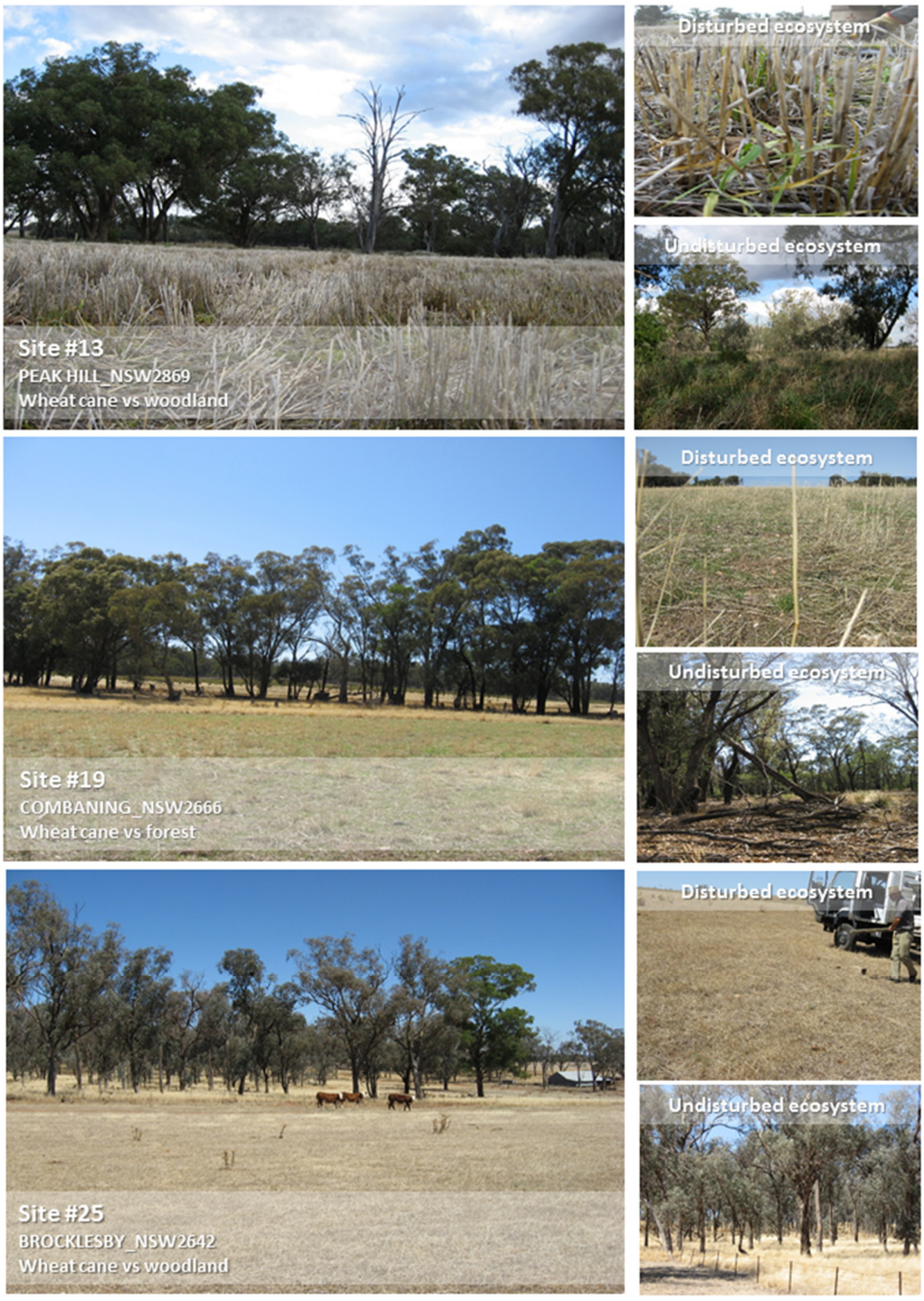


Figure 2-34. Some of the sampling sites sampled showing differences between undisturbed and disturbed ecosystems sampled in the north of NS-$_{transect}$.

Figure 2−35. Some of the sampling sites sampled showing differences between undisturbed and disturbed ecosystems sampled in the north of WE$_{-transect}$.

ENVIRONMENTAL GRADIENT ALONG THE TRANSECTS

The main environmental gradients included in our analyses are display in the maps below in the context of our entire study area. We can infer from these maps that $WE_{\text{-transect}}$ involves more environmental variations than the $NS_{\text{-transect}}$. In Figure 2-36, for example, we can see from an ecological perspective that $WE_{\text{-transect}}$ includes nearly five different types of bioclimates (from Mediterranean Xeric to Temperate Hyperoceanic) and nearly six different categories of agroclimatic zones (from Dry to Subtropical Moist). Whereas that $NS_{\text{-transect}}$ comprises only two type of bioclimates (Temperate Semi-continental and Mediterranean Pluviseasonal) and mostly three categories of agroclimatic zones (Subtropical Sub-humid, Temperate Sub-humid and Temperate Cool-season Wet).

More precisely, the most remarkable environmental changes seen from west to east along $WE_{\text{-transect}}$ are the increases in both precipitation and elevation patterns, and a decrease pattern in temperatures. Towards the east, precipitation pattern ranges between 170 – 3,200 mm/year and the elevation can range between 0 - 2,100, meanwhile, temperatures decrease from 21 to 10 °C (Figure 2-36).

Regarding soil change, some defined gradients seen from west to east in $WE_{\text{-transect}}$ are the decreases in both soil pH and plant available water capacity (PAWC). In the first 30 cm, soil pH ranges from 6-7 in the west and goes down to 4-5 in the east part of the transect. In the case of

PAWC, the first 100 cm of the top soil can range from 290–110 mm of water capacity in the extreme west to less than 80-60 mm of water capacity when passing through the elevated regions of the Eastern Uplands (Figure 2-37). Soil clay content is variable along the transect. For the first 30 cm depth, the percentage of clay content shows to increase from 30% in the extreme west of the transect to about 70% in the area of Darling Riverine Plain, being the maximum values just in middle of the transect. From these areas to the east, the soil clay content starts to descend to reach less than 20-30% in the zones nearby the coastline. Similarly, bulk density (Db) fluctuates up and down along the $WE_{\text{-transect}}$ from being very low to up to values of 1.7 Mg/cm^3 but scarcely distributed in few areas; however, in general terms it seems that Db tends to be lower towards the eastern side.

On the other hand, the $NS_{\text{-transect}}$ involves only those areas where the precipitation pattern is about 550 mm/year. Likewise, elevation does not present drastic fluctuations along the transect by keeping most of its values ranging between 180-320 m and only a bordering few areas up to 540-570 m in the southern part. Likely, it is the reduction of temperature from north to south one of the most significant environmental gradients in this transect, which oscillation gradually ranges between 21°C and 15 °C.

The soil gradient along the $NS_{\text{-transect}}$ shows a reduction pattern from north to south in some of the soil properties. Soil pH, for example, has values between 8 and 9 in the areas located in the northern part, while towards the southern areas it ranges between 6 and 7. Clay content presents an important decline by fluctuating from 70% in the northerly zones to 10% in some locations

of the south. Soil plant water availability capacity in the north can range around 290-110 mm in the first 100 cm depth and, on the way down to the south it can be found ranges of 60-66 mm. Conversely, bulk density (Db) is one of the properties that increase from north to south, ranging from being even less than 1.0 Mg m^{-3} in the north to reach nearly 1,7 Mg/m^3 in many of the southern areas (Figure 2-37).

In the next chapter…

The next chapter 3 describes the protocols applied for microbial and soil physicochemical analyses carried out using on the soil samples extracted from the 49 sampling sites composing $NS_{-transect}$ and $WE_{-transect}$.

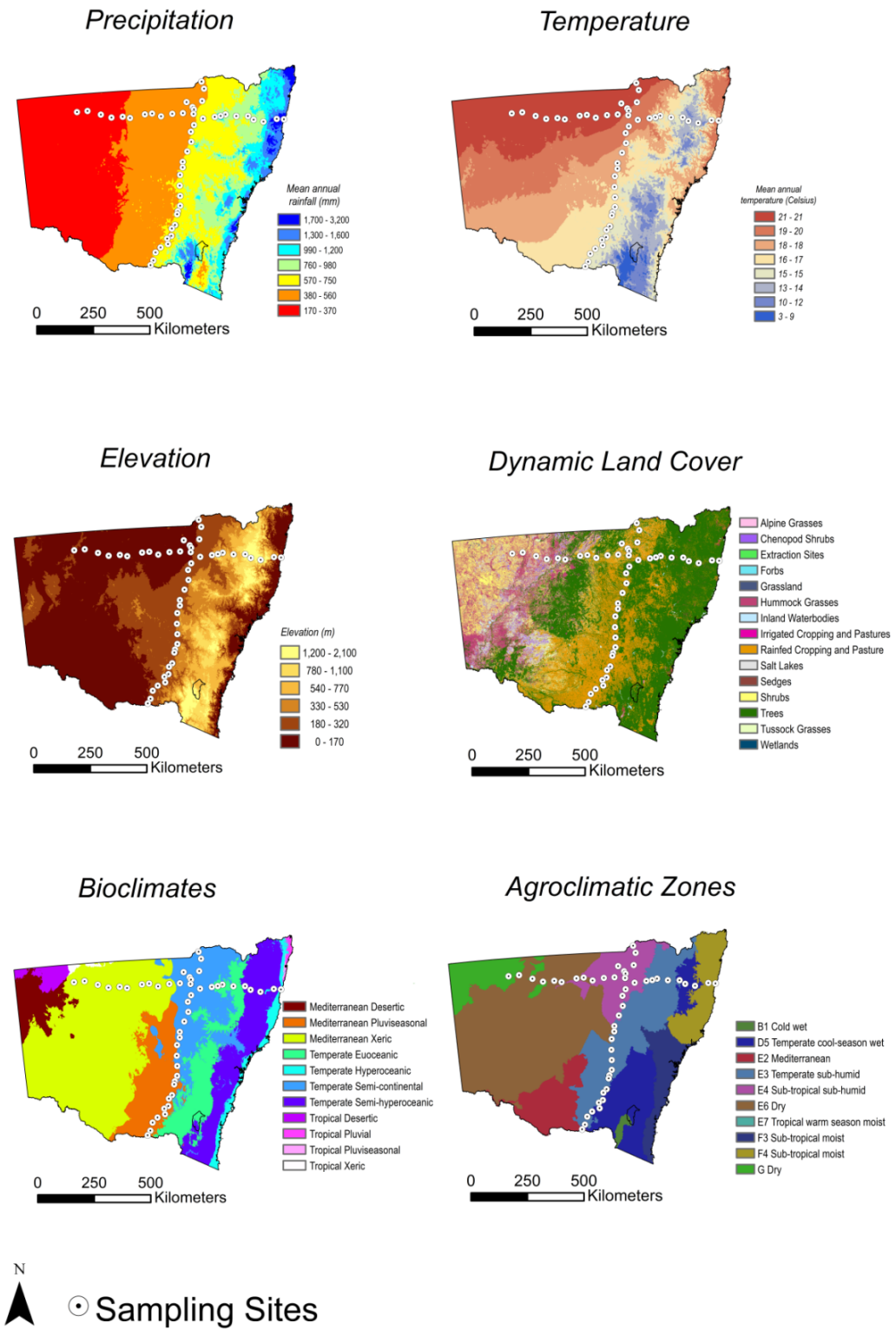


Figure 2-36. NSW environmental gradients (ABARES, 2014).

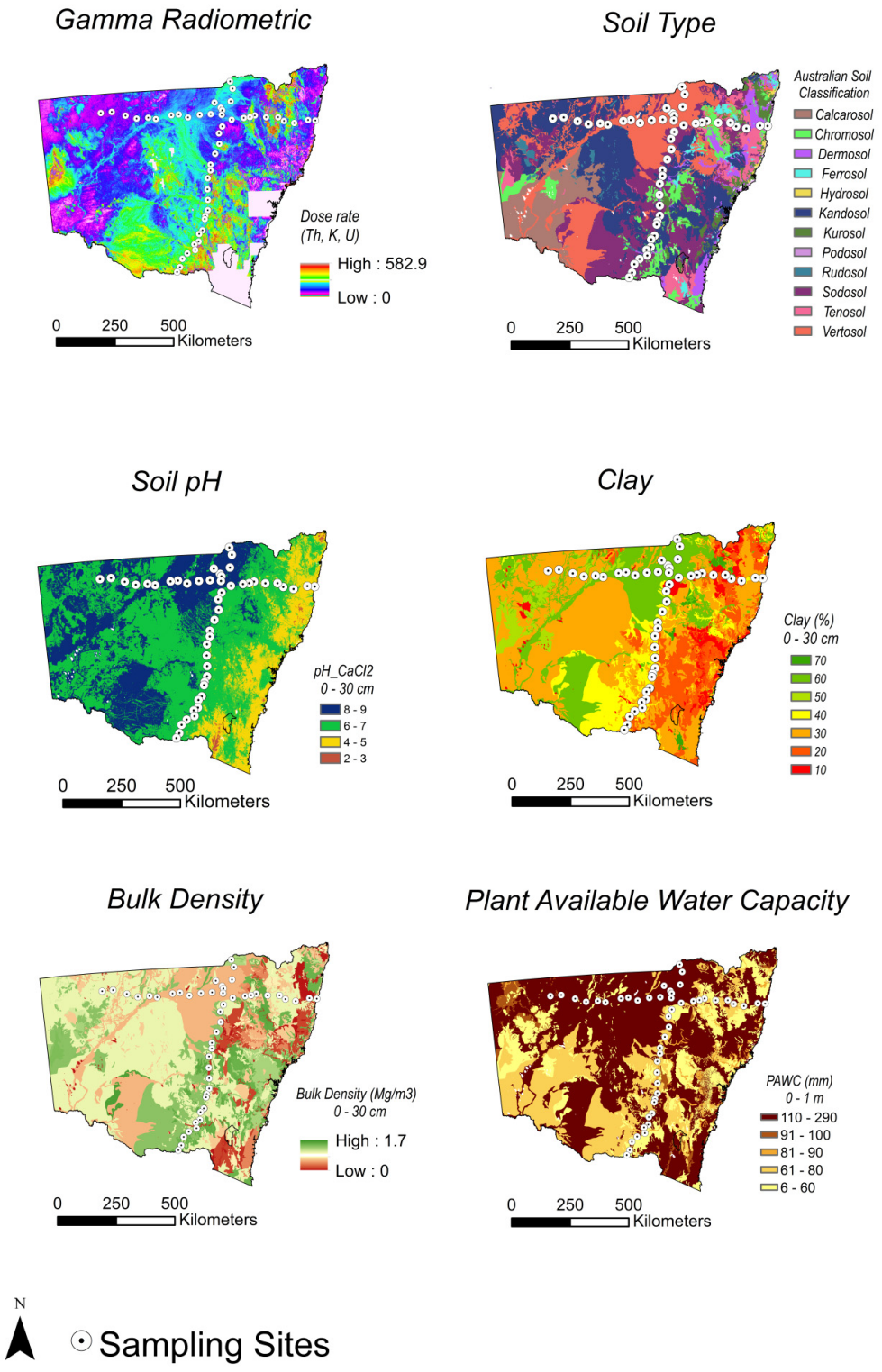


Figure 2-37. NSW soil gradients (ABARES, 2014).

REFERENCES

ABARES, 2014. 2014 Datapack to support the Multi-Criteria Analysis Shell for Spatial Decision Support (MCAS-S).

ABARES, 2012. Economics and Sciences, Agricultural Commodities Statistics 2012.

ABS, 2015a. Murray-Darling Basin, in: Pink, B. (Ed.), Year Book Australia, 2009–10, Year Book of the Commonwealth. Australian Bureau of Statistics, Canberra, Australia, p. 861.

ABS, 2015b. Land Management and Farming in Australia, 2013-14, Agriculture Statistics. Australia.

ABS, 2013. Agricultural Commodities, 2011-12, Agriculture Statistics, Agricultural Commodities. Australia.

ACLEP, 2011. The Physiographic Regions of Australia. Australian Collaborative Land Evaluation Program and National Committee on Soil and Terrain, Canberra, Australia.

Australia Bureau of Meteorology (BOM), 2009. Mean monthly and mean annual rainfall data (base climatological data sets) [WWW Document]. URL http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp (accessed 11.7.12).

Australia Bureau of Meteorology (BOM), 2005. Climate classification maps [WWW Document]. URL http://www.bom.gov.au/jsp/ncc/climate_averages/climate-classifications/index.jsp?maptype=kpngrp#maps (accessed 3.1.16).

Australian Government, D. of E., 2001. Biodiversity conservation research: Australia's priorities [WWW Document]. URL

http://www.environment.gov.au/node/14372#ecosystems (accessed 2.24.16).

Blewett, R., 2012. Shaping a nation : a geology of Australia, 34th International Geological Congress.

CAPAD, 2014. The Collaborative Australian Protected Area Database [WWW Document]. Aust. Gov. Dep. Enviroment. URL http://www.environment.gov.au/land/nrs/science/capad (accessed 3.22.16).

Charman, P.E. V, Murphy, B.W., 1991. Soils, their properties and management: A soil conservation handbook for New South Wales. Sydney University Press in association with Oxford University Press, South Melbourne, Australia.

CSIRO, 2001. Conversion of the Atlas of Australian Soils to the Australian Soil Classification.

DSEWPaC, 2012a. Interim Biogeographic Regionalisation for Australia [WWW Document]. URL http://www.environment.gov.au/land/nrs/science/ibra (accessed 4.15.15).

DSEWPaC, 2012b. Australia's ecoregions [WWW Document]. Dep. Environ. Natl. Reserv. Syst. URL http://secure.environment.gov.au/parks/nrs/science/bioregion-framework/terrestrial-habitats.html (accessed 3.24.16).

EPA, 2012. NSW State of the Environment 2012. Sydney Environ. Prot. Auth. NSW 302.

ESRI, R., 2011. ArcGIS desktop: release 10. Environ. Syst. Res. Inst. release 10.

GA, 2010. The National Dynamic Land Cover Dataset. Geoscience Australia, Canberra, Australia.

GADDS, 2010. The Radiometric Map of Australia Dataset. Geoscience Australia, Canberra.

Hill, S.M., 2004. Regolith and landscape evolution of far western New South Wales. Regolith Landsc. Evol. across Aust. 1–17.

McKenzie, N., Jacquier, D., Isbell, R., Brown, K., 2004. Australian Soils and Landscapes: An

Illustrated Compendium. Csiro Publishing.

Mckenzie, N.J., Jacquier, D.W., Ashton, L.J., Cresswell, H.P., 2000. Estimation of soil properties using the Atlas of Australian Soils. Database.

MDBC, 1999. Groundwater a Resource for the Future. Murray-Darling Basin Commission, Canberra.

Morton, S., Lonsdale, M., Sheppard, A., CSIRO, 2014. Biodiversity : science and solutions for Australia, Biodiversity : science and solutions for Australia. CSIRO Publishing.

NARCliM, 2014. New South Wales Climate change snapshot Overview of New South Wales climate change [WWW Document]. URL http://climatechange.environment.nsw.gov.au/Climate-projections-for-NSW/Climate-projections-for-your-region/NSW-Climate-Change-Downloads (accessed 4.16.15).

NRMMC, 2010. Australia's Biodiversity Conservation Strategy 2010–2030. Australian Government, Department of Sustainability, Environment, Water, Population and Communities, Canberra.

Chapter 3

Soil Microbial & Physicochemical Measurements

Outlines

Soil physical, chemical and microbiological properties were characterized along both transects using modern 'biodiversity and pedometrics' measurement approaches. The microbial characterisation was reached by evaluating the structural diversity of the major soil microbial taxa (i.e. bacteria, archaea, and fungi) by identifying them and estimating afterwards their presence/abundance and phylogenetic composition through alpha and beta diversity indices. Simultaneously, soil physicochemical properties were characterised by both conventional laboratory as well as other pedometrics measurement techniques. Soil microbial diversity and physiochemical data are related in depth in the next chapter. In this chapter are described the laboratory protocols by which the soil microbial and physicochemical datasets were generated.

Summary

Soil microbial and physicochemical diversities were evaluated on the samples derived from the forty-nine sampling sites composing both $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$. To conduct this microbial and physiochemical characterisation at a high-scale resolution both soil microbial taxonomy and physiochemical measurements have been obtained using modern biodiversity and pedometrics approaches.

Soil microbial taxonomic identification has been based on a high-throughput sequencing DNA gene-based method. In this method, the strategy applied start with DNA extraction, then amplification for 16S rDNA and ITS genes with barcoded primers. These barcoded amplicons were sequenced with the Illumina MiSeq system. The sequencing data derived was processed using a series of open pipelines available in QIIME and Pipits sources for the respective identification-quantification of *bacteria/archaea* and *fungi* OTUs. Thus, the entire protocol involved with microbial identification and abundance estimation includes all the laboratory procedures from soil DNA extraction until picking up, identifying the OTUs and evaluating their abundance within each soil sample.

Concurrently, the soil physiochemical properties have been generated using both (i) conventional laboratory analyses and (ii) pedometrics techniques (*e.g.* infrared spectroscopy). Some of the soil physicochemical properties measured by conventional methods included Total Carbon (TC), Extractable Phosphorus (P), Effective Cation Exchange Capacity (ECEC), Electric Conductivity (EC), pH in water and particle size (*e.g.* clay content). Other soil properties and

characteristics of the soil composition were analysed on the basis of Vis-NIR, mid-IR instruments and image processing recognition. Some of the properties estimated by these pedometrics techniques were slaking index (SI), to estimate soil structural stability; and soil spectral horizonation to provide some morphologic features of the soil profiles to 100 cm depth (*e.g.* number of horizons).

Further in Chapter 4, the soil microbial taxonomic dataset based on OTUs is used to calculate microbial alpha and beta diversity to evaluate the structure of microbial communities. Subsequently, both soil microbial and soil physicochemical properties are related to each other in the context of other environmental variables (*e.g.* rainfall, temperature, etc.) characterising $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$.

INTRODUCING THE SOIL MICROBIAL AND PHYSICOCHEMICAL CHARACTERISATION

To provide evidence of spatial soil biodiversity-pedodiversity relationship, soil microbial biogeographical patterns by analysing the microbial structural diversity (biodiversity) has been assessed in relation with the soil physicochemical variation (pedodiversity) and other key environmental variables.

The soil microbial characterisation was made analysing the structural diversity of *fungi*, *bacteria* and *archaea* communities and others *unassigned prokaryotes*. This microbial structural diversity has been established based on (i) the microbial taxonomic identification-quantification and, subsequent (ii) alpha and beta diversity analyses to assess richness, evenness, similarities and dissimilarities of the microbial communities which will be presented in Chapters 4 and 5. Figure 3-1 briefly describes that microbial taxonomic characterisation has been obtained following four main steps: (1) soil DNA template preparation, (2) dual-barcoded target genes template preparation, (3) IIlumina 16S rDNA/ITS amplicon sequencing and (4) OTUs identity-quantification. This entire procedure led to decoding the molecular genetics information of the soil microorganisms using a high-throughput sequencing technology, which is nowadays one of the most revolutionary for microbial identification (Nesme et al., 2016; Torsvik L., 2002; Zarraonaindia et al., 2013). In this process, soil DNA extraction and amplification of the target genes (step 1 and 2) implied all the laboratory routines to prepare the soil DNA template for the following DNA processing on the sequencing platform. Up to this point all the laboratory

procedures were carried out in facilities of the Faculty of Agriculture and Environment of the University of Sydney in Sydney, Australia. Illumina sequencing performance (step 3) was performed by Micromon Laboratory in the Department of Microbiology of the Monash University in Victoria, Australia. Then, the sequencing Illumina output data - *i.e.* ~ 47,115,100 sequences (~31.6 GB data) – were treated with a set of bioinformatics tools to decode and assign the corresponding taxonomic information (*e.g.* pick OTUs) for the different taxonomic groups (step 4).

Soil physicochemical characteristics were assessed by conventional laboratory analyses of key properties most commonly reported as influencing soil microbial communities, *e.g.* carbon, nitrogen, pH, electrical conductivity, among others (step 1). Moreover, other additional information on soil variation such as the soil spectral absorbance measurements has been provided using other pedometrics techniques (step 2). The resulting ranges dataset derived from this measurements are shown at the end of this chapter (step3).

Microbial and soil physicochemical properties relationship is analysed in Chapter 4 according with (1) their co-spatial changes along the two transects and (b) their specific parallels and interactions with specific properties and environmental factors that might be influencing the structural diversity of these microbes. This information would be useful to assess how closely related are biodiversity-pedodiversity in soil ecosystems. Below are described the technical methods followed to generate the microbial identification-quantification and soil physicochemical measurement datasets.

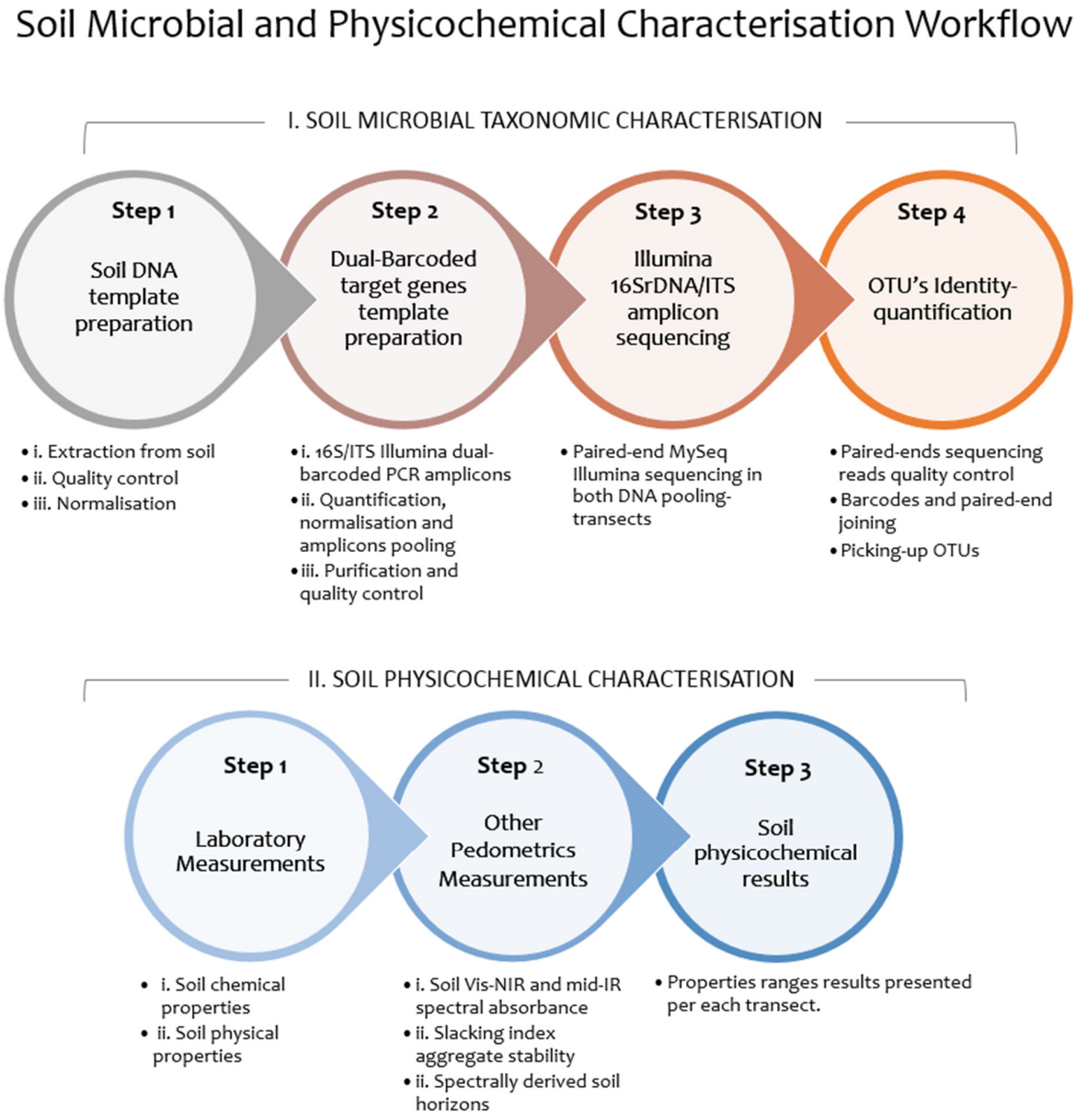


Figure 3−1. General workflow followed to obtain microbial and soil physicochemical properties databases.

SOIL MICROBIAL TAXONOMIC CHARACTERISATION

Soil microbial taxonomy was obtained based on a DNA gene sequencing strategy on Illumina sequencing platform. This entire workflow involved all the procedures from the soil DNA extraction in laboratory conditions until the final microbial identification when analysing *in silico* the DNA sequences reads derived from the Illumina sequencing. These DNA sequences were decoded to obtain the identity and abundance of microorganisms on the basis of the recommended minimum taxonomic level of sampling when working with DNA sequencing approaches, *i.e.* the operational taxonomic unit (OTUs) (Blaxter et al., 2005; Torsvik et al., 1998).

Microbial identification involved the entire procedure as far as the quantification of the operational taxonomic units (OTUs) present per each of the soil samples. In other words, the soil microbial taxonomy was obtained during this process by decoding the DNA sequences contained in the molecules of the soil DNA. These soil DNA sequences were obtained using a *DNA metabarcoding sequencing approach*, *i.e. using metabarcoding marker genes on 'a priori' targeted DNA region* (Taberlet et al., 2012). This method correspond to a *culture- independent strategy* for microbial characterisation that does not involve microbial cultivation.

More technically, the target microbial DNA sequences were obtained from the small subunit ribosomal DNA (SSU rDNA) per each taxa, *i.e.* 16S rDNA genes of *bacteria/archaea* and the *Internal transcribed spacer (ITS)* genes of *fungi*. These genes were amplified and indexed by mean

of the 'eDNA[xxi] metabarcoding PCR-based' method (Mendoza et al., 2014; Taberlet et al., 2012)– also referred as *amplicon sequencing method* (Fierer et al., 2012; Larsen et al., 2012; Nesme et al., 2016; Sinclair et al., 2015) or *amplicon metagenomics* (Xu et al., 2015; Zarraonaindia et al., 2013).

In other words, this meant that the microbial target soil DNA fragments were amplified to up to billions of copies and then indexed by a dual-barcoding preparation in a two-step polymerase chain reaction (PCR). Finally, the dual-barcoded PCR amplicons, derived from the two group of genes (*i.e.*16S rDNA and ITS) were pooled into one aliquot. This pooling of amplicons was finally sequenced by performing a paired-end sequencing protocol on the Illumina MiSeq platform. Once obtained the Illumina dataset, the soil microbial communities were identified and quantified using available open-source pipelines (QUIIME, Pipits, others) which were set up following the developer's recommendations according to the nature of the microbial data generated to be used in subsequent diversity analyses (*e.g.* .biom statistics) (Caporaso et al., 2010b).

This DNA metabarcoding sequencing approach can generally be described into four main stages:

- STEP 1. SOIL DNA TEMPLATE PREPARATION
- STEP 2. DUAL-BARCODED TARGET GENES TEMPLATE PREPARATION
- STEP 3. AMPLICON ILLUMINA SEQUENCING

[xxi] *Total eDNA involves cellular DNA (living cells or organisms) and extracellular DNA (resulting from natural cell death and subsequent destruction of cell structure). Soil DNA contains 'intracellular DNA' mainly from bacteria, fungi, roots and 'extracellular DNA' from all organisms living around (bacteria, fungi, plants, animals, etc.)* (Levy-Booth et al., 2007).

- STEP 4. MICROBIAL IDENTITY-QUANTIFICATION

Below are described step by step each of these laboratory, sequencing and bioinformatics procedures together with the technical set up and conditions used in the context of this research. Every downstream laboratory routine was performed under sterile conditions. Likewise, soil samples and their derived genetics products were carefully manipulated and kept in benchtop coolers at ~2°C and immediately stored at -20 °C until subsequent steps.

STEP 1. SOIL DNA TEMPLATE PREPARATION

This procedure comprised the preparation of the soil DNA product that was afterwards used as template for PCR1 routine, when amplifying the target genes for microbial identification. The laboratory method involved in this procedure included: (i) isolation (ii) quantification/quality control and, (iii) normalisation of these soil DNA templates. Each of these activities are detailed below.

(i) Soil DNA extraction

Soil DNA was extracted from each sample using the PowerSoil DNA isolation Kit from MO BIO Laboratories, Inc. This commercial kit bases the isolation of the genomic material on combining mechanical, physical and chemical processes to lyse the cells and release the DNA by using the advantage of the Inhibitor Removal Technology (IRT). IRT increases the effectiveness of removing inhibitors (*e.g.* humic and fulvic acid) for downstream procedures (*e.g.* PCR amplification) and also facilitates the deletion of the tight binding between DNA strands and clay particle (MO BIO Laboratories, 2014). For this reason, this has become a very useful method when extracting DNA from the most challenging type of soils such as the Vertosols (clayey soil) which are found in our study area. Additional advantages of this commercial kit are the precision and speed when handling a considerable number of samples.

DNA isolation Kit protocol

Soil DNA was extracted from 0.25g of soil on the basis of the PowerSoil DNA isolation Kit manufacturer's protocol (MO BIO Laboratories, 2014). The illustration below (Figure 3-2) shows the workflow followed to proceed the cells lysis phase until the DNA is release and captured on a silica membrane from where it is finally purified, washed and eluted to obtain 100 µl of DNA extract per each sample. A total of 588 soil DNA samples from 0-10 cm depth were processed and finally stored at -20°C in a freezer at the Faculty of Agriculture and Environment of The University of Sydney.

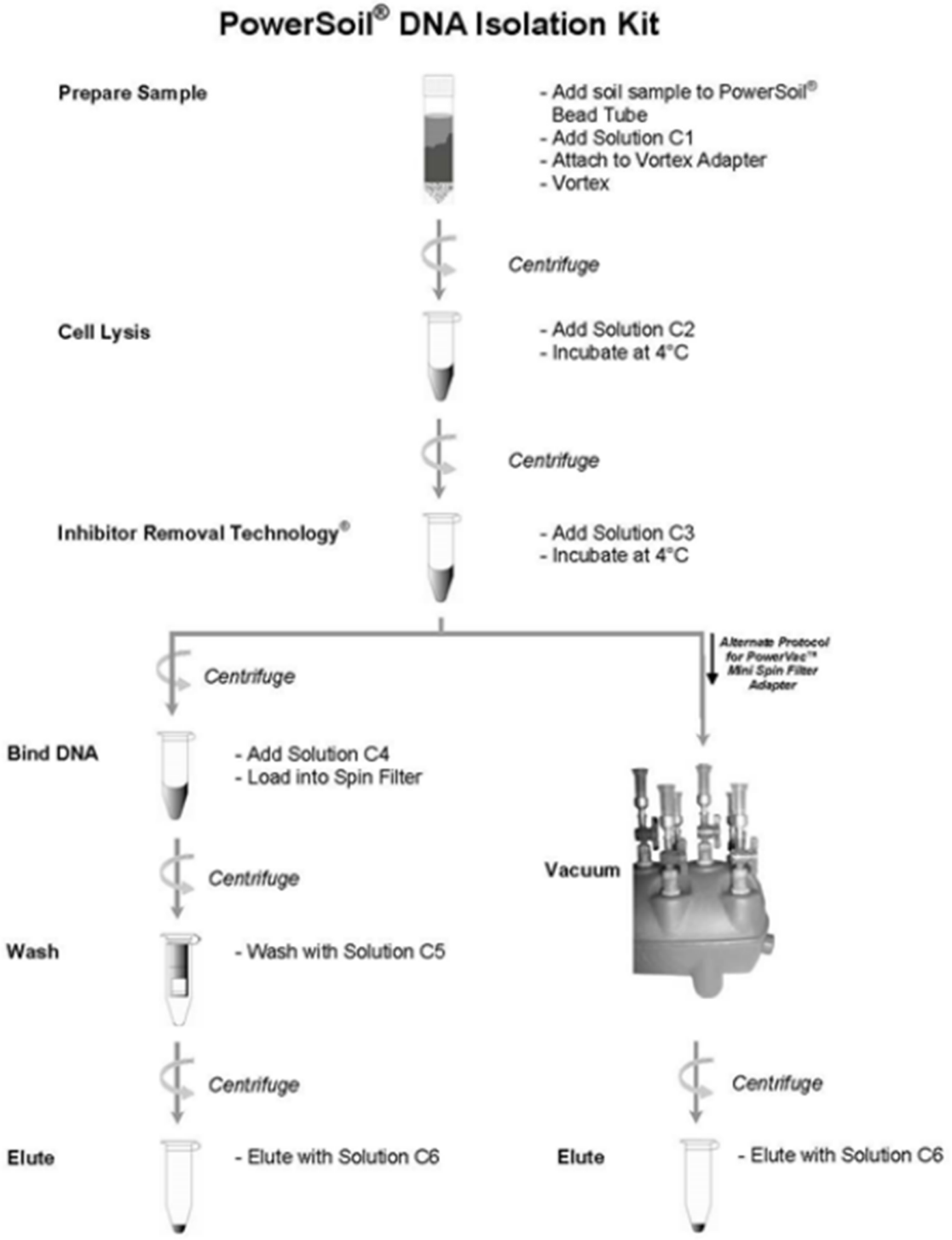


Figure 3-2. PowerSoil DNA isolation Kit flowchart for DNA extraction (MO BIO Laboratories, Inc.). In our extractions we followed the centrifuge method on the basis of moist soil.

(ii) DNA quantitation: quality control

Immediately after being extracted, the soil DNA samples were analysed to verify their quality, quantity and molecular weight in term of the nucleic-acid concentration (ng/µl). This information was necessary for DNA normalisation in downstream procedures. Since there are various techniques available for measuring nucleic-acid concentration which range various advantages and disadvantages when working with environmental DNA - soil samples are the most challenging for gene-based methods- we have implemented three types of them:

a) Agarose gel horizontal electrophoresis
b) UV-Vis spectrophotometry
c) Fluorimetry

d) Agarose gel electrophoresis by bromide ethidium

This is a common laboratory techniques to detect, quantify and purify nucleic acid fragments according to their molecular size (Aaij and Borst, 1972). In this procedure, the DNA samples are loaded into wells of an agarose gel and passed through an electrical field in which the nucleic acid are attracted to the positive electrode due to its negative charge (Figure 3-3). As result, since the shorter DNA fragments travel faster and they end clearly separated from the longest fragments, it is possible to estimate their size, concentration and molecular weight using standard molecular weight markers (Southern, 1979).

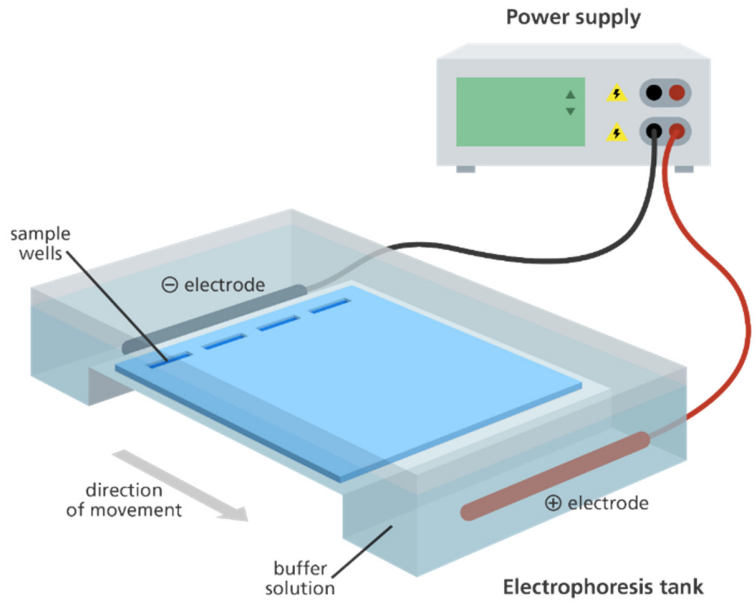


Figure 3−3. Electrophoresis Equipment. Image credit: Genome Research Limited.

In this case, an ethidium bromide agarose gel at 1% concentration using Tris-acetate-EDTA (TAE) as buffer solution has been prepared. Once the gel solidified, 10µl of DNA product were added into the gel wells for each sample. Next to the DNA samples, a molecular marker well was added with 5µl of Bioline HyperLadder 1kb to determine molecular sizes in the range of 200 bp to 10,037bp (20 to 100 ng/band)[xxii]. Each gel electrophoresis was run at 84 volts for ~20 minutes. The resulting DNA bands were optimized and captured under UV-LED light illumination using a ChemiDoc™ MP Imaging System[xxiii] (Bio-Rad Laboratories, Inc.). The soil DNA images coming from the samples taken at 0-5cm depth are shown organized for each sampling site below in Figure 3-7.

[xxii] *Manufacturer's product description in* http://www.bioline.com/au/hyperladder-1kb.html.

[xxiii] *Manufacturer's User Guide in* http://www.bio-rad.com/webroot/web/pdf/lsr/literature/10022469.pdf

e) UV-Vis spectroscopy by NanoDrop™ 2000

DNA nucleic acid concentration has also been measured on the basis of UV-Vis spectrophotometer technology using NanoDrop™ 2000 (pedestal mode) manufactured by Thermo Scientific Inc. The equipment operates producing a surface tension to place the sample in between two optical fibres which is able to measure absorbance between wavelength of 190 nm and 840 nm. In the nucleic acid module, DNA sample purity was calculated based on the 260/280 nm and 260/230 nm ratios. This nucleic acid calculations are based on a modification of the Beer-Lambert equation which results are reported in ng/μL. The upper and lower detection limits in the case of the DNA double-stranded (dsDNA) can range between 2 and ≤ 15,000 ng/μL.

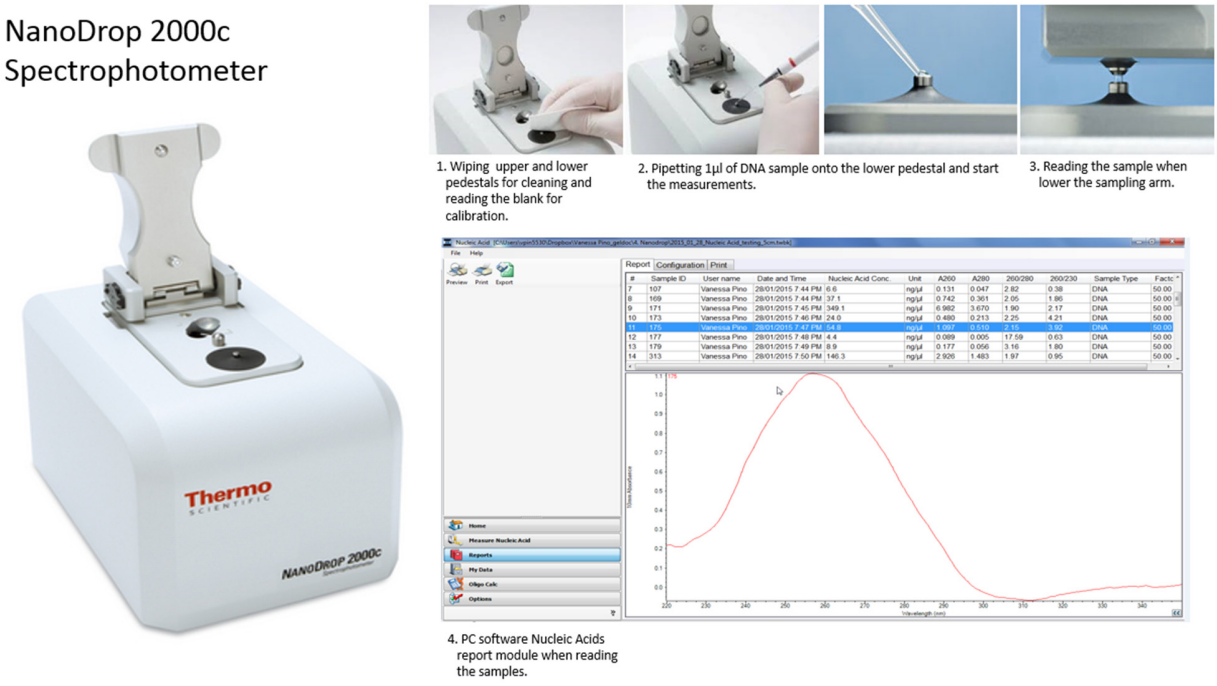


Figure 3–4. UV–Vis spectrophotometer equipment for nucleic acid concentration measurements. Image credits: Thermo Scientific Inc.

Our soil DNA samples were quantified using a micro-volume of no more than 1µl per sample using the 'nucleic acid application' module of the equipment. A previous blank absorbance calibration was made periodically using the same solution buffer (10 mM Tris) used for the final DNA elution step when the DNA was extracted from the soil sample (C6 solution from PowerSoil DNA isolation Kit).

f) Fluorometry by Quant-iT™ PicoGreen®dsDNA Assay Kit

Genomic DNA was also analysed using a fluorescence-based quantitation method. This method required pre-treating the soil DNA sample with a fluorescent nucleic acid stain followed by measurement by a florescence detector. For this method, the DNA samples were pre-treated using a Quant-iT™ PicoGreen®dsDNA Assay Kit from Thermo Scientific Inc. The fluorescent reagent that used in this assay (*i.e.* Hoechst-bisbenzimide dye), binds to nucleic acids in the range 502-523 nm, one which it has orders of magnitude more sensitivity than UV absorbance readings at 260 nm (A_{260}). This allows quantification of very low concentrations of dsDNA by detecting even ranges of about 0.01 ng/µl (Figure 3-5). This method is very selective with dsDNA upon ssDNA, RNA and others free nucleotides present in the samples (Figure 3-5) and therefore, it also increases the accuracy of readings for dsDNA[xxiv].

First, the Quant-iT™ PicoGreen dsDNA assay requires preparation of a series of purified Lambda DNA standard solutions by diluting the provided DNA stock 100 µg/mL with the TE buffer diluent (10 mM Tris - HCl, 1 mM EDTA, pH 7.5). Accordingly to both the provider's specifications

[xxiv] *Full version of Quant-iT™ PicoGreen® manufacturer's description* https://tools.thermofisher.com/content/sfs/manuals/mp07581.pdf

and empirical proofs for soil dsDNA samples, it was prepared a high-range standard curve based on four standards stocks ranging final concentrations between 10 to 500 pg/µl (0.01 – 0.5 ng/µl). This is a very crucial step since accuracy in readings by these methods rely on the quality of the standard curve.

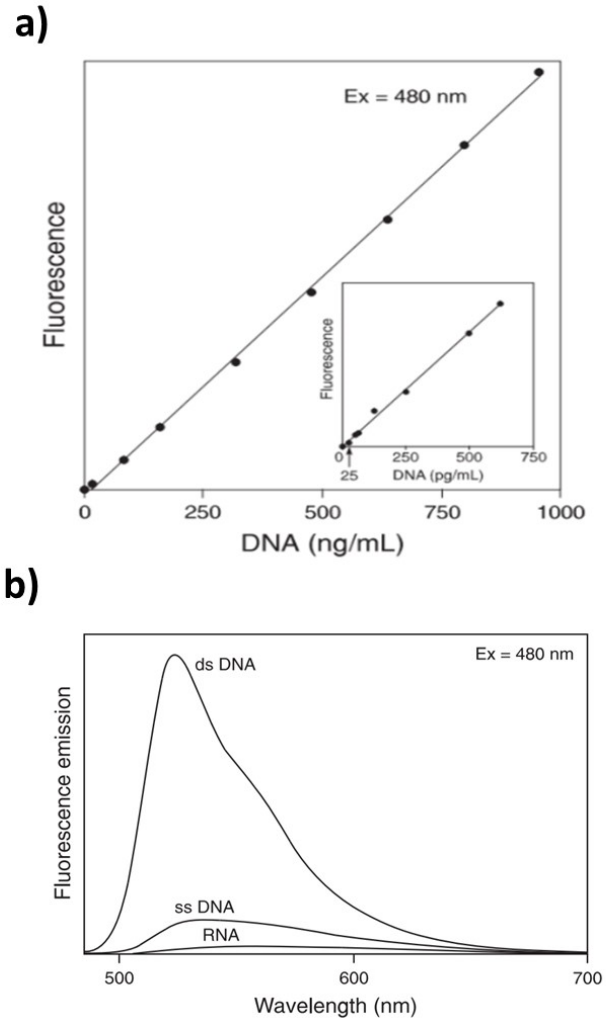


Figure 3–5. a) Fluorescence enhancement of Quant–iT™ PicoGreen®. In this case is shown the emission spectra for samples containing dye and nucleic acid, as well as for dye alone (baseline). b) Dynamic range and sensitivity of the Quant–iT™ PicoGreen® dsDNA assay. It is shown the fluorescence emission intensity plotted versus DNA concentration; the inset shows an enlargement of the results obtained with DNA concentrations between zero and 750 pg/mL.

Secondly, the DNA samples to be quantified require to be diluted because of the sensitivity of this method and its detection limits. Therefore, we diluted 2µl of each soil DNA sample in 98µl of the TE buffer (10 mM Tris - HCl, 1 mM EDTA, pH 7.5). Finally, we added 100µl of the Picogreen reagent on the top of 100µl of each of the diluted DNA samples and 100µl of each of the DNA standards to enhance their fluorescence. Also few control blanks containing only TE buffer were included. These solutions were incubated for 5 minutes in the dark at room temperature afterwards the samples were ready to be quantify.

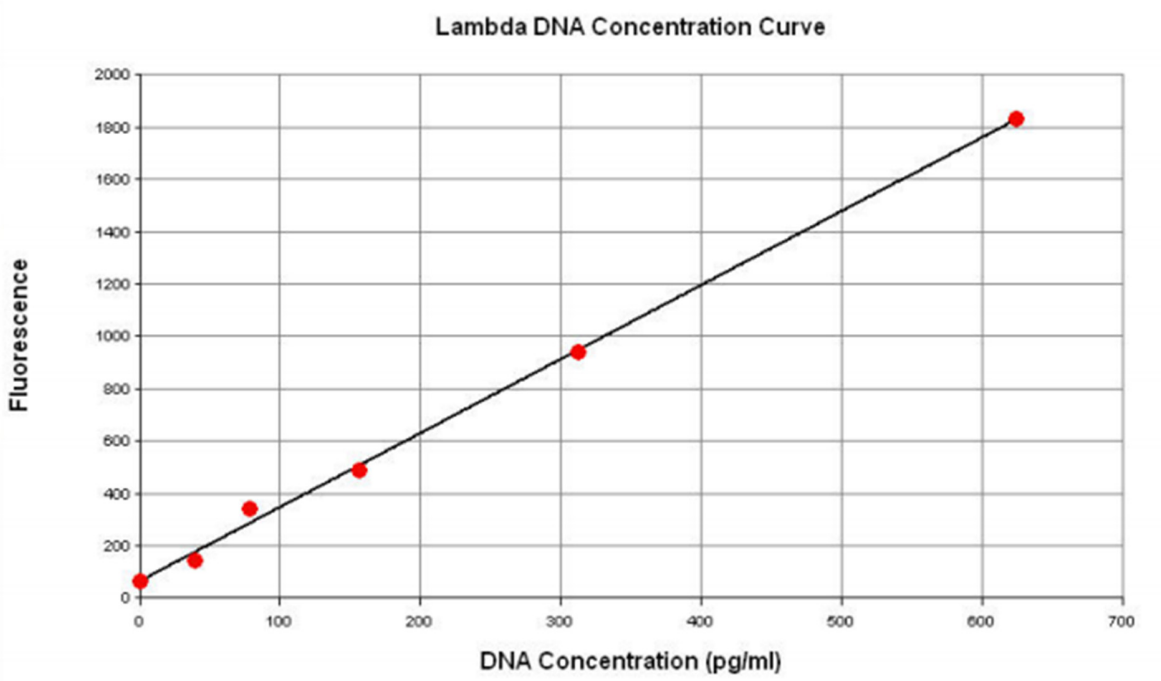


Figure 3-6. Fluorescence of Lambda dsDNA concentrations from 0-1000 ng/ml. Example provided by the BioTek manufacturers to illustrate the performance of the Gen5™ Data Analysis Software.

The final DNA concentration was estimated by measuring the fluorescent signal from all the samples using the Synergy™ H4 Hybrid Multi-Mode Microplate Reader from BioTek Instrument, Inc.

The final readings were run using Gen5™ Data Analysis Software (BioTek Instruments, Inc.) in which the samples were excited at 480 nm and the fluorescence emission intensity was measured at 520 nm by the spectrofluorometer. On the basis of the standard solutions signals, the software generated a least means squared linear regression analysis to construct the standard concentration curve[xxv] as exemplified in the Figure 3-6 above.

DNA quantitation results

Table 3-1 below shows the ranges of DNA concentration (ng/µl) values obtained in soil DNA samples from 5 cm depth and quantified by Quant-iT™ PicoGreen®dsDNA Assay. These ranges of values are displayed for each transect and ecosystem of the study area. These results are additionally accompanied by the soil DNA electrophoresis outputs illustrated in the Figure 3-7. Electrophoresis output showing soil DNA bands from $NS_{\text{-transect}}$. 10 µl loaded. MM HyperLadder, Bioline®.. As explained above (on page 161), these images display the DNA bands that provides a qualitative estimation of the DNA concentration per each individual sample. Together these and other results (e.g. many from NanoDrop) have been used to decide in the next procedure either to dilute highly concentrated samples or re-extract low concentrated samples to normalise

[xxv] http://www.biotek.com/resources/docs/Fluorometric_Quant_dsDNA_PicoGreen_AppNote.pdf

them preceding the PCR's routines (i.e. gene amplification).

Table 3−1. DNA concentration ranges values obtained per each transect and ecosystem.

DNA concentrations (ng/μl)	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
			$NS_{-transect}$			
Natural ecosystems	2.0	9.0	12.9	17.8	20.5	90.0
Cropping ecosystems	2.0	8.5	12.5	19.0	19.1	85.0
			$WE_{-transect}$			
Natural ecosystems	5.8	11.8	17.9	20.6	27.1	58.1
Cropping ecosystems	1.0	7.8	15.7	19.6	29.5	62.6

(iii) Soil DNA template normalisation

It is commonly recognized that DNA concentration might largely affect sequencing readings in downstream analyses. Accordingly, it is recommended to standardise these values as much as possible to a common concentration (Kennedy et al., 2014). Hence, preceding the PCR amplification of the target 16S rDNA and ITS genes, all the samples composing the soil DNA stock were normalised to ranges between 5 to 10 ng/μl. These ranging values were defined running several optimisation tests for PCR1 condition on which also different DNA template dilutions were tested (*e.g.* 1:1; 1:10; 1:100, etc.). The criteria applied to define such ranges was to optimize the DNA template concentrations ranges in order to reduce as much as possible the inclusion of methodological biases (Kennedy et al., 2014). The most of our samples worked better normalised up to 1:10 dilution.

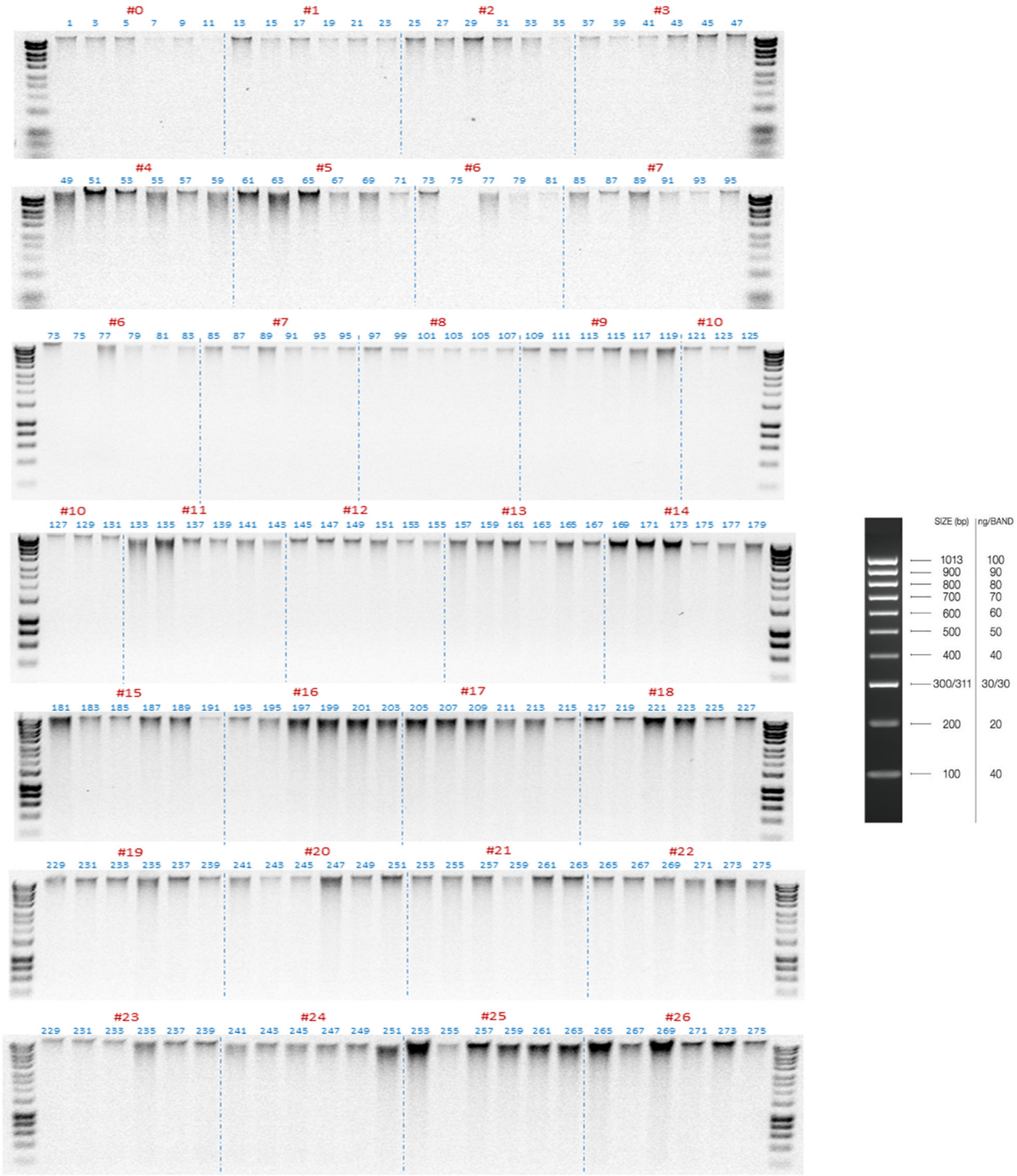


Figure 3–7. Electrophoresis output showing soil DNA bands from $NS_{-transect}$. 10 µl loaded. MM HyperLadder, Bioline®.

The final volumes required for each soil DNA sample were calculated to store a final normalised DNA volume between 15 to 20µl. This calculation was made using the concentration measurements derived from Quant-iT™ PicoGreen®dsDNA Assay, which provided more accurate ranges than NanoDrop Assay. Highly concentrated samples at >10 ng/µl were diluted using 5mM Tris - HCl pH 8 solution. Low concentration samples at <5 ng/µl were either re-extracted or purified using the Isolate II PCR and Gel Kit Bioline® manufacture's protocol. It is noteworthy that only soil samples belonging to site #46 from $WE_{\text{-transect}}$ were re-extracted and purified twice due to low concentration yield.

Output from step1…

Step 1 has been concluded with a set of 294 DNA templates samples distributed into four 96-multiwell plates comprising an entire normalised DNA library with two plates per each transect stored at -20°C.

STEP 2. DUAL-BARCODED TARGET GENES TEMPLATE PREPARATION

This step includes all the laboratory processing required to prepare the microbial DNA template to be sequenced on the Illumina instrument. As specified above, microbial identification has been based on a DNA metabarcoding PCR-based approach for which 16S rDNA and ITS regions were barcoded (Mendoza et al., 2014; Taberlet et al., 2012). This means that the sequencing reads were made over the DNA amplicons (*i.e.* replicates of DNA) obtained from these DNA regions. According to this principle and aiming to ensure fidelity and depth in the sequencing reads, a paired-end sequencing performance on MySeq platform has been selected, which ensures production of 250-nucleotide paired reads (Kozich et al., 2013). Under a scheme of paired-end sequencing both forward and reverse terminals of the DNA fragments replicated were sequenced at the same time. As a result it generates a large number of high-quality sequences which must produce a highly precise alignment of the reads and improves the accuracy in the microbial identification results[xxvi]. On top of that, the procedure was quite accessible since, it did not require costs associated to the use of long customized primers (Fadrosh et al., 2014; Kozich et al., 2013).

To aim a paired-end sequencing strategy on the Illumina platform, the microbial DNA amplicon 16S rDNA and ITS templates were prepared following a dual-barcoded PCR amplicon protocol

[xxvi] *http://www.illumina.com/technology/next-generation-sequencing/paired-end-sequencing_assay.html*

adapted from the Earth Microbiome Project[xxvii] according to a series of recent published methods. Then, the final protocol was set up to firstly obtain and secondly to condition the DNA amplicon to be sequenced by the Illumina platform. Briefly, this dual-barcoded laboratory protocol involved the following activities:

i. Generation of dual-barcoded PCR amplicons library by
    A) PCR1: amplification of target 16S rDNA and ITS
    B) PCR2: Illumina barcoding of 16S rDNA and ITS amplicons

ii. Quantification and pooling of PCR amplicons library

iii. Amplicon pools purification and quality control

The specific laboratory conditions applied per each of these routines are described below.

(i) Generation of dual-barcoded PCR amplicons library

Largely, *bacteria/archaea* and *fungi* are identified *via* 16Sr DNA and ITS genes, respectively. Then, most of DNA-sequencing-based analyses such as 'DNA metabarcoding approaches' works on sequencing these DNA regions to recognise presence of these taxa. Likewise, the distinguishing of these target genes by a DNA metabarcoding approach can be either be by sequencing them directly and randomly, *i.e.* using a shotgun sequencing strategy or by amplifying thousands replicates fragments of them (amplicons) using specific molecular markers (*i.e.* primers), the strategy applied in this research.

[xxvii] *Sample processing, sequencing and core amplicon data analysis were performed by a modified pipeline of the Earth Microbiome Project (www.earthmicrobiome.org).*

Since we opted for paired-end sequencing strategy in the MySeq Illumina platform, the main purposes when generating the amplicon libraries were both the amplification of 16Sr DNA/ITS genes and the addition of the Illumina barcodes and adapters through a two-step PCR routines. A schematic view of this procedure is illustrated in the Figure 3-8 and is further described below.

Both PCR1 and PCR2 routines were carefully optimised according to the technical requirements of the reagents used (*e.g.* optimal temperatures for DNA polymerase *Phusion* enzyme and molecular primers) and, also in accordance with the optimal quality of the PCR products required for downstream analyses. This latter was directly related to the number of PCR cycles and the template concentration used in every PCR reaction, *i.e.* the normalised DNA for PCR1 and PCR1 amplicon product for PCR2, for example.

The reiterative recommendations found in the literature about the imminent methodological bias that might be included when running these PCRs routines has been considered a quite important criteria for the PCR's optimisations. For this reason, the number of PCR cycles was reduced as much as possible in defining the final protocols for 16S rDNA and ITS amplification. Such protocols as well as the specification of PCR conditions are presented below together with other details related to each of the PCR's performances. The enzyme supplied in both PCR's was Phusion® High-Fidelity DNA Polymerase (New England Biolabs® Inc.) which is certified as a highly thermostable enzyme[xxviii].

[xxviii] *With an error rate > 50-fold lower than that of Taq DNA Polymerase and 6-fold lower than that of Pyrococcus furiosus DNA Polymerase (1)* https://www.neb.com/products/m0530-phusion-high-fidelity-dna-polymerase

PCR1: amplification of target 16S rDNA and ITS genes

PCR1 was the first reaction performed on the soil DNA template already normalised. This first PCR reaction aimed: (i) to produce 16S rDNA and ITS amplicons containing both (ii) heterogeneity spacers for improving sequences diversity, and (iii) the universal PCR primer-specific adapters that allowed to enable the addition of the barcodes in PCR2. Schematically, the above exemplified for the 16S rDNA amplicon library is illustrated in Figure 3-9.

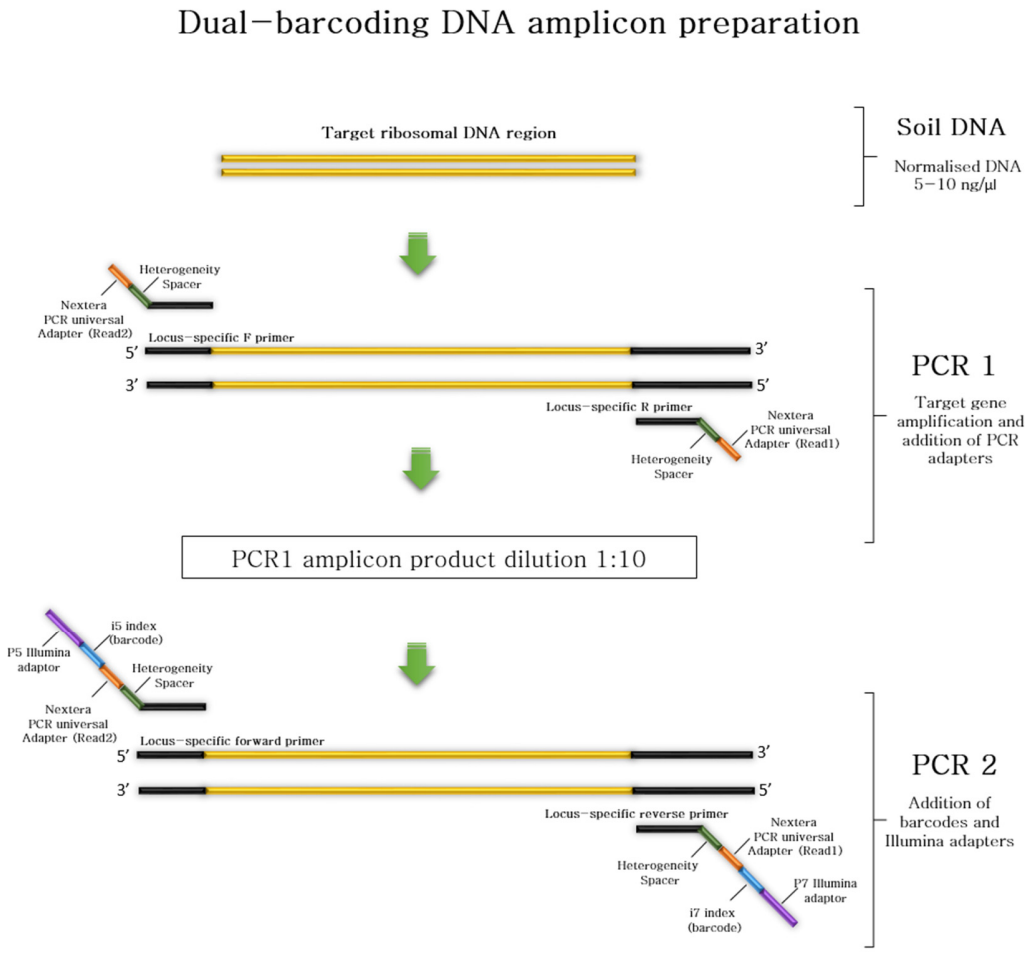


Figure 3−8. Schematic view of the DNA amplicon library generation during the two−step PCR reaction routines.

a) PCR1 primers library for target genes amplification

16S rDNA and ITS amplicons have been amplified using molecular markers which have been reported in the literature as being of high-coverage when used by a dual-barcoding approach (Takahashi et al., 2014; Toju et al., 2012). On the basis of these molecular markers, *bacteria/archaea* (prokaryotes) were simultaneously amplified from V3-V4 hypervariable regions of the prokaryotic rDNA by using a mix of Pro341F/805R primers that allow production of fragments of about ~400 base pairs (Figure 3-10) (Takahashi et al., 2014). On its own, *fungi* taxa were amplified from the ITS2 spacer region of the eukaryotic rDNA using a mix of ITS3_KYO2-F/ITS4-R primers which produces fragments of less than ~700 base pairs (Figure 3-11) (Toju et al., 2012).

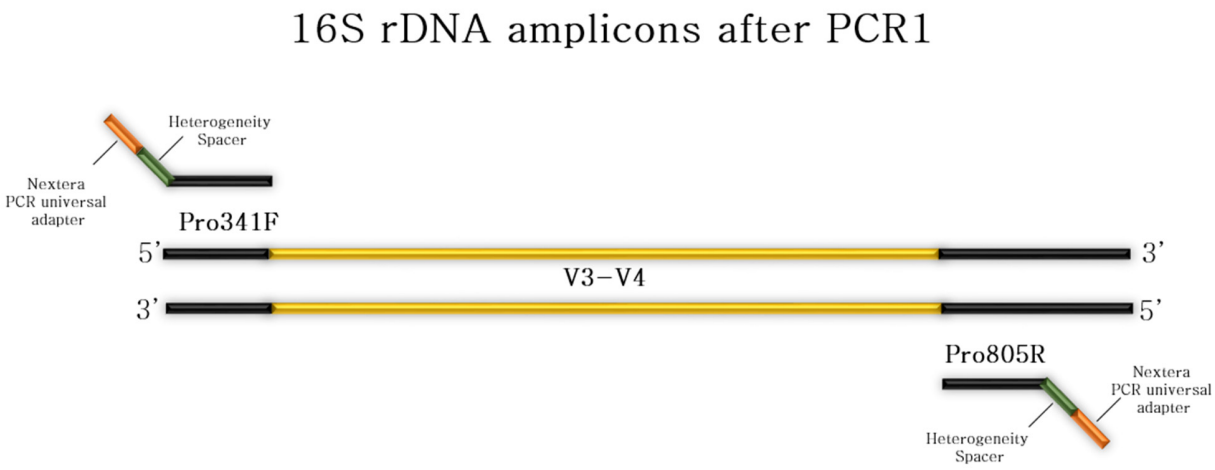


Figure 3–9. Schematic view to 16S rDNA amplicons Illumina barcoding after PCR1 reaction.

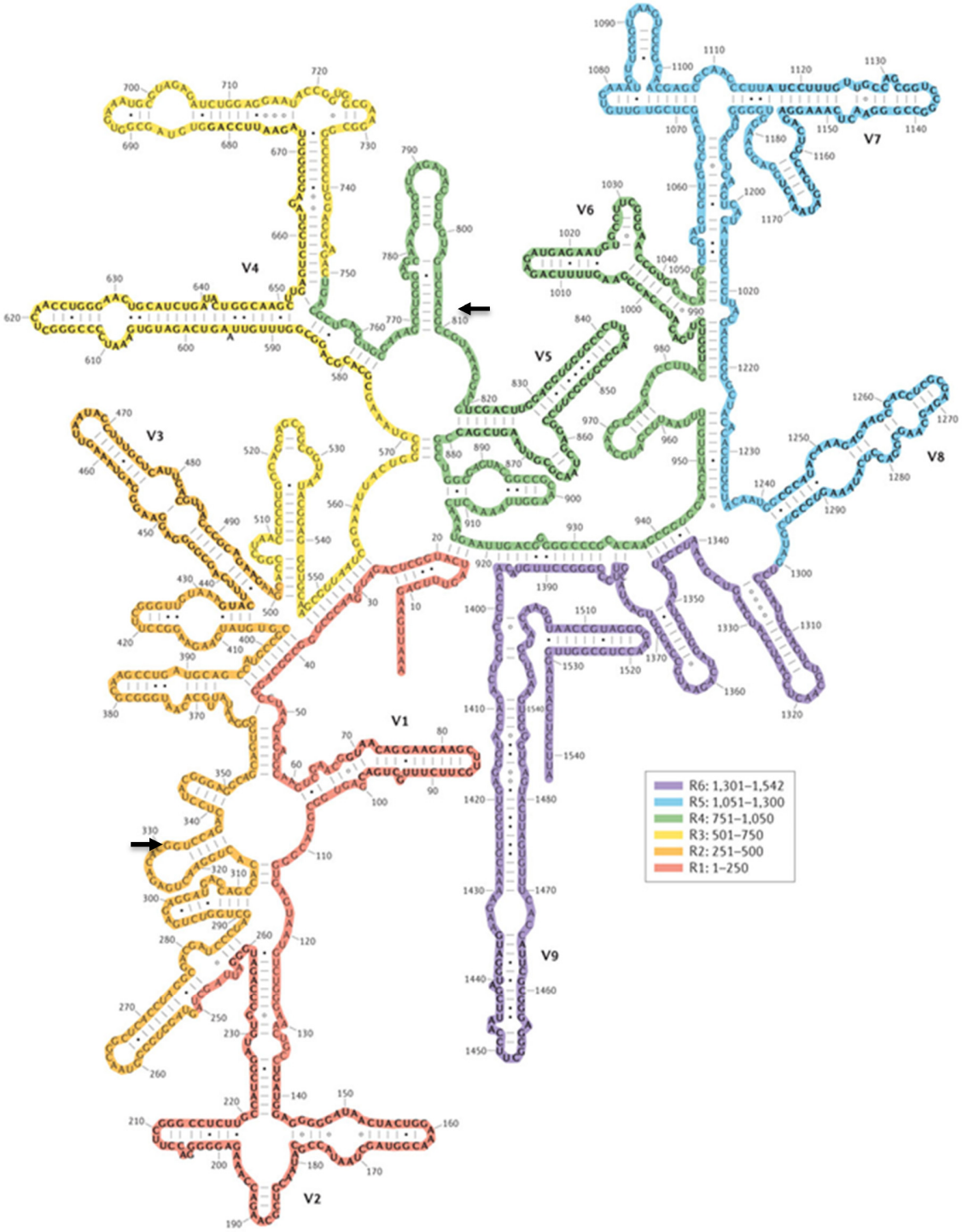


Figure 3−10. Secondary structure of the 16S rDNA of *Escherichia coli*. In this survey, V3−V4 hypervariable regions were sequenced using Pro341F/Pro805R for Prokaryotes primers. This image reference was taken from Yarza et al., (2014).

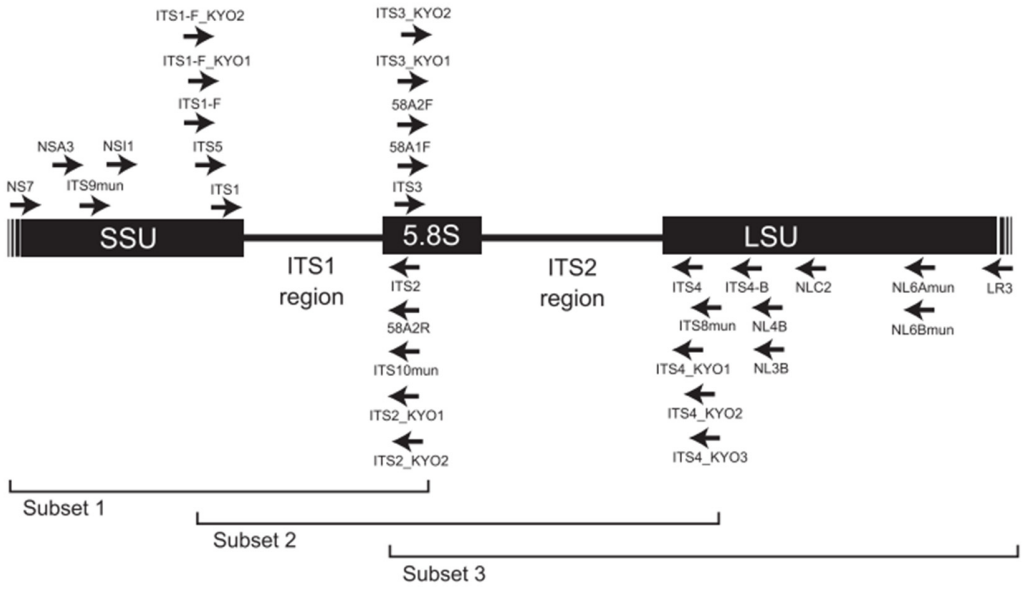


Figure 3−11. Map of nuclear ribosomal RNA genes and their ITS regions as shown by **Toju et al., 2012**. ITS2 was the target region amplified in this survey by using ITS3_KYO2−F and ITS4−R primers.

b) <u>Improving sequencing quality by the insertion of heterogeneity spacers:</u>

To improve sequencing quality results, both forward and reverse sets of 16S rDNA (V3-V4) and ITS2 spacer primers, have been designed in combination with 'heterogeneity spacers sequences'. As published by Fadrosh et al., (2014), the use of these 'heterogeneity spacers' are advantageous by reducing the difficulties encountered when sequencing samples with low sequence diversity on the Illumina platform. In this case, the primer mixes have been designed using 0 (without), 2, 4 and 6 base pair lengths as heterogeneity spacers; the specific sequences of which are shown in the Table 3-2.

Table 3−2. Primer mixes sequencing scheme used in PCR1 reaction. It is shown 16S rDNA and ITS primers mixed with heterogeneity spacers and Illumina adapters.

16S rDNA primers mixes	
	Forward
Pro341F_H0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTACGGGNBGCASCAG
Pro341F_H2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCTACGGGNBGCASCAG
Pro341F_H4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGGCCTACGGGNBGCASCAG
Pro341F_H6	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATGTCCTACGGGNBGCASCAG
	Reverse
Pro805R_H0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTACNVGGGTATCTAATCC
Pro805R_H2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGACTACNVGGGTATCTAATCC
Pro805R_H4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTACGACTACNVGGGTATCTAATCC
Pro805R_H6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGTTGTGACTACNVGGGTATCTAATCC
ITS primers mixes	
	Forward
ITS3_KYO2_H0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCGATGAAGAACGYAGYRAA
ITS3_KYO2_H2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTCGATGAAGAACGYAGYRAA
ITS3_KYO2_H4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAGTCGATGAAGAACGYAGYRAA
ITS3_KYO2_H6	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTATTCGATGAAGAACGYAGYRAA
	Reverse
ITS4_H0	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTCCGCTTATTGATATGC
ITS4_H2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCCTCCGCTTATTGATATGC
ITS4_H4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGGTCCTCCGCTTATTGATATGC
ITS4_H6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTAAGGTCCTCCGCTTATTGATATGC

* Underlined sequences (34 bases) indicate forwards (Read2_GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) and reverse (Read1_TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) Illumina adapters Nextera® transposase sequences. ** Highlighted sequences correspond to the heterogeneity spacers used per primer mix combination (*i.e.* 2, 4 and 6 bases).

c) Adding universal PCR adapters:

During PCR1 universal PCR adapters were added, these are the so-called 'Nextera transposase sequences'. These forward and reverse adapters are essential for enabling the subsequent addition of barcode sequences (*i.e.* i5 and i7 index) in the following PCR2 routine (Figure 3-8). Since this gene amplification process has been treated as a dual-barcoded preparation, both

forward and reverse 16S rDNA and ITS amplicons ended carrying the universal adapters. In this case, we used Nextera transposases 34 base-pairs large, the sequence of which are specified the Table 3-2.

d) Preparing PCR1 primers mix library:

According to the scheme above in Figure 3-8, both 16S rDNA and ITS forward and reverse specific-primers (*i.e.* Pro341F/Pro805R and ITS3_KYO2/ITS4) in combination with heterogeneity spacers and Nextera adapters led to the generation of a PCR1 primer library of 16 primer mixes (Table 3-2). Then, forward and reverse primers mixes were combined into four different primer mix 'combinations' (1-I4) (Table 3-3). These primer combinations were used to prepare the 10 µl master solution for PCR1 reaction. To improve sequence diversity amplicon, an equal number of DNA samples were prepared using each of these libraries into the 96-multiwell plates.

Table 3−3. Primer mix combinations used to prepare both 16S rDNA and ITS PCR1 reactions.

Forward and reverse 16 rDNA (V3−V4) primer mix combination		Forward and reverse ITS2 primer mix combination	
Name	Primers	Name	Primers
16S_I1	Pro341F_H0 + Pro805R_H6	ITS_I1	ITS_KYO2_H0 + ITS4_H6
16S_I2	Pro341F_H2 + Pro805R_H4	ITS_I2	ITS_KYO2_H2 + ITS4_H4
16S_I3	Pro341F_H4 + Pro805R_H2	ITS_I3	ITS_KYO2_H4 + ITS4_H2
16S_I4	Pro341F_H6 + Pro805R_H0	ITS_I4	ITS_KYO2_H6 + ITS4_H0

* Each of this primer mix library was equally used across the DNA samples.

e) Preparing PCR1 master solution reaction:

PCR1 has been run on 10 µl solution reaction in both 16S rDNA and ITS reactions. The master solution was prepared using 2µl of normalised DNA and other reagents specified in the Table 3-4 below. In total, 294 samples reactions were prepared for this run which were distributed in sets of eight 96-multiwell plates, four plates for each $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$. For each of these sets we also included negative controls as well as positive control (mock community) for the *16S rDNA* plates. As mentioned above, each of the primers-mix combinations specified in Table 3-3 were equally distributed per each 96-multiwell plates as shown schematized in the Figure 3-12.

Table 3−4. PCR1 master mix volumes and concentration reagents used for 10 µl solution reaction.

Reagent	Per 10 µl sol. reaction	Final conc.
MQ H_2O	4.7 µl	
5 X Buffer	2 µl	1 X
10mM dNTPs	0.2 µl	200 µM
PCR 1 primer mix (5 µM)	1 µl	0.5 µM
Phusion Polymerase	0.1 µl	0.2 units
Template DNA	2 µl	10−20 ng

f) PCR1 cycling condition:

PCR1 cycling conditions were optimised according to the polymerase enzyme (*Phusion*) and the specific set of primers (Pro341F/805R and ITS3_KYO2/ITS4) requirements. Then, the final cycling temperatures for denaturation, annealing and extension reactions during PCR1 were programmed in a BIO-RAD C1000 Touch™ Thermal Cycler applying the conditions in Table 3-5.

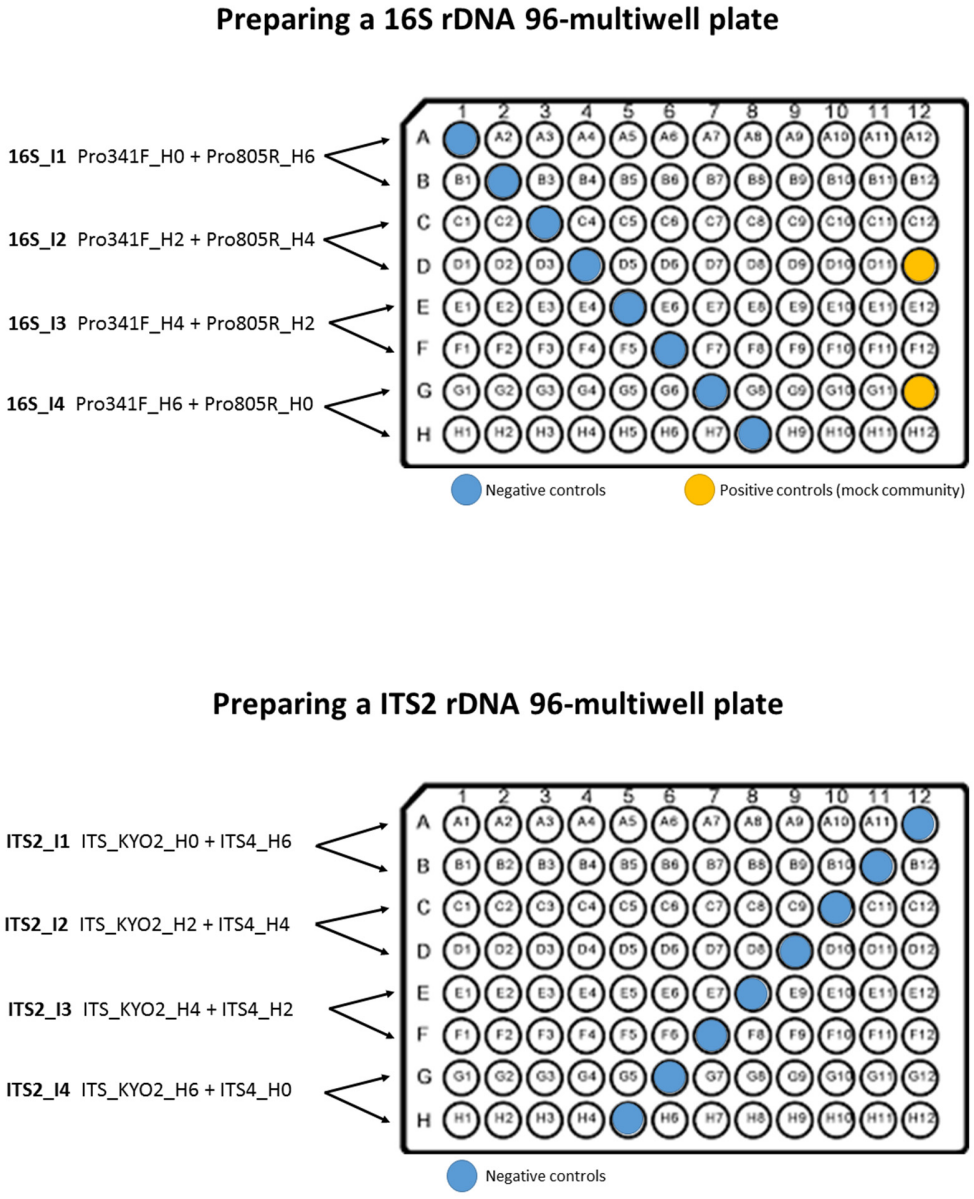


Figure 3-12. Schematic procedure when preparing PCR1 16S rDNA and ITS rDNA 10 μl solution reaction into 96-multiwell plates.

Table 3−5. Optimal thermo−cycling conditions for PCR1routine using Phusion enzyme.

Reaction	Temperature	Time	Cycles
Initial Denaturation	98° C	1:30	X 1
Denature	98° C	0:10	
Annealing	55° C	0:20	X 25
Extension	72° C	1:00	
Final Extension	72° C	10:00	X 1
	8° C	Hold	

* PCR thermos cycler was warmed up to 98 °C previous to place the samples.

g) PCR1 product quantification:

PCR1 amplicon products were quantified during different testing runs using the same methods applied when quantifying DNA templates concentrations specified above (*i.e.* gel electrophoresis, Quant-IT Picogreen, etc.). However, in the final routine, the 96-well plates were directly processed through the PCR2. An important precaution taken was to manipulate the PCR1 product in another room when preparing PCR2 reactions to reduce contamination.

PCR2: Illumina barcoding of 16S rDNA and ITS amplicons

A second PCR reaction was performed using as template PCR1 product diluted 1:10. This second PCR routine aimed to add in both gene-specific PCR1 products: (i) the barcodes sequences for indexing each sample and, (ii) the Illumina sequencing adapters which allow Illumina readings. Schematically, this is exemplified for 16S rDNA dual-barcoding amplicon library in the Figure 3-13 below.

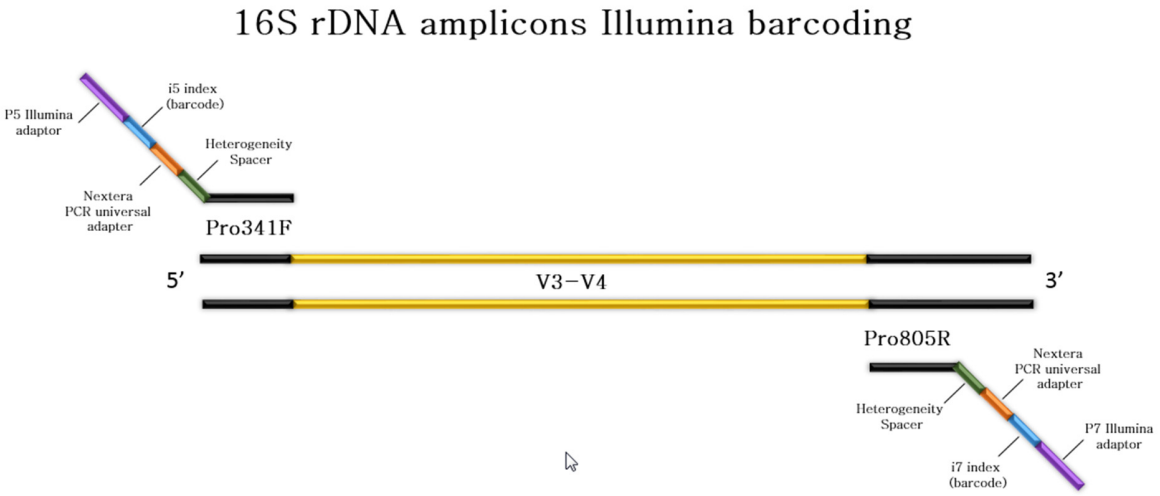


Figure 3−13. Schematic view to 16S rDNA amplicons Illumina barcoding after PCR2 reaction.

h) PCR2 primers library for amplicon barcoding:

PCR2 primers were designed into a mix that contained the barcoding sequences (index) and the Illumina adapters. First, index barcodes allowed a sample multiplexing for one direct reading on the Illumina platform. These indices permitted mixing of all the samples when pooling their PCR2 products from both 16S rDNA and ITS genes into one pool for Illumina readings – producing one pool per each transect. The barcode library was prepared combining a set of 16 different forward index sequences (i5) and 12 different reverse index sequences (i7) of 8 bp length each. These together generated a library of 384 index combinations into a set of four 96-multiwell plates. The same barcode library was used for each transect. Secondly, the Illumina adapters were the linker sequences allowing amplicons to bind the Illumina flow cell to be sequenced.

i) Preparing PCR2 master solution reaction:

PCR2 was run on 20 µl solution reaction in both 16S rDNA and ITS amplicon libraries. In this reaction has been used 2 µl of 1:10 diluted PCR1 product. This dilution was made adding 90 µl of 5mM Tris - HCl pH 8 directly to the 10 µl of PCR1 product into the 96-multiwell plates. Other PCR reagents were added as indicated in Table 3-6 below. In the same way as PCR1, 294 samples together with negative and positive controls were processed in PCR2 using the eight sets of 96-multiwell plates with four plates per each transect.

Table 3−6. PCR2 master mix volumes and concentration reagents used for 20 µl solution reaction.

Reagent	Per 10 µl sol. reaction	Final conc.
MQ H_2O	3.6 µl	
5 X HF Buffer	4 µl	1 X
10mM dNTPs	0.2 µl	200 µM
Phusion Polymerase	0.2 µl	0.5 units
Template PCR1 (1:10)	2 µl	~variable per sample

j) PCR2 cycling conditions:

PCR2 thermo-cycling conditions were optimised according to the reagent requirements and the resulting PCR2 products on which were determined sample quality for pooling and sequencing. PCR2 routines were run using the same thermos-cycler instrument under the conditions specified below (Table 3-7). PCR2 amplicon products were kept stored at -21°C.

Table 3−7. Optimal thermo−cycling conditions for PCR2 routine using Phusion enzyme.

Reaction	Temperature	Time	Cycles
Initial Denaturation	98° C	1:00	X 1
Denature	98° C	0:10	X 12 (16S)* X 15 (ITS)*
*Annealing	55° C	0:20	
Extension	72° C	0:45	
Final Extension	72° C	10:00	X 1
	8° C	Hold	

* Numbers of cycles were obtained from different optimisation tests. ** PCR thermos cycler was warmed up to 98 °C previous to place the samples.

(ii) Quantification and pooling of PCR amplicons library

k) DNA amplicon quantification and normalisation

To verify that PCR2 worked out positively, a subset of 20 samples for each 96-multiwell plate, positive and negative controls included, were qualified by electrophoresis on agarose gel. These gels roughly showed that we obtained 16S rDNA bands ranging ~600 bp and ITS bands ranging ~ 520 bp. These results provided the licence to continue to Quant-iT™ PicoGreen quantification of all the samples composing the amplicon library. DNA amplicon quantification was necessary to estimate the volume (µl) required for each sample for the subsequent pooling procedure. Those samples that either failed to show electrophoresis bands or which the concentrations values were less than 4 ng/µl and so would require to exceed 10 µl volume for being pooled, were repeated or concentrated. In total, no more than 12 samples were repeated and/or concentrated using the ethanol precipitation method.

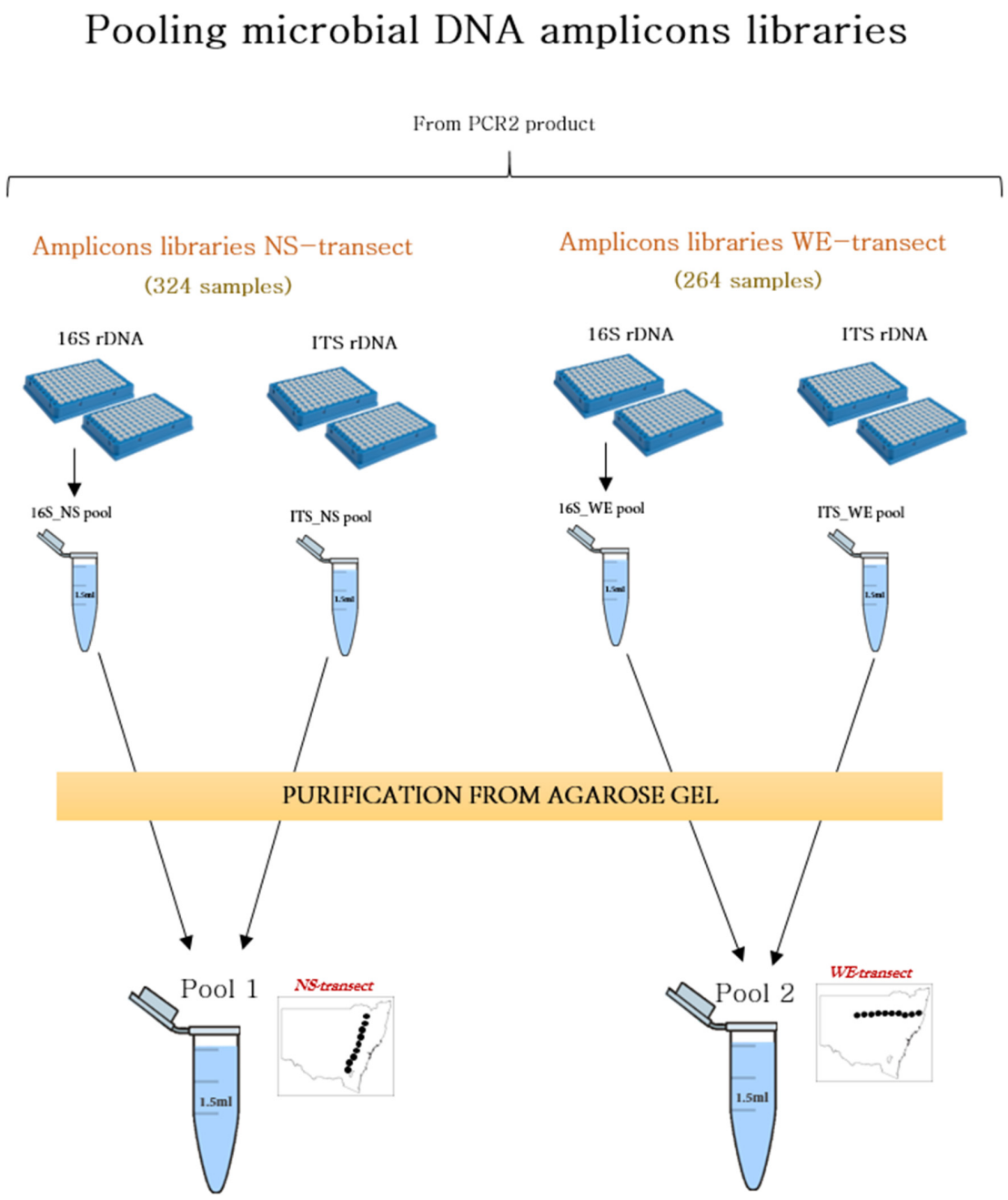


Figure 3−14. Pooling 16S rDNA and ITS amplicons from PCR2 products in accordance with Illumina MySeq requirements for samples preparation.

l) Pooling the DNA amplicon library

All the DNA amplicon samples composing 16S rDNA and ITS libraries for each $NS_{-transect}$ (162 samples) and $WE_{-transect}$ (132 samples) were pooled using <10 µl per sample and concentrating of about 40-160 ng each into one single tube - this is possible since each sample per sampling site, replicates included, were carefully barcoded in PCR2 - as describe on page 183. Thus, these libraries derived from the 96-multiwell plates into two 1.5 ml tubes as it is illustrated in the Figure 3-14 above.

(iii) Amplicon pools purification and quality control

m) Amplicon pools purification

16S rDNA and ITS microbial DNA amplicon pools were purified using a gel clean- up method. This method required extraction of the DNA product directly from an agarose gel immediately after the electrophoresis run. According to this, 50 – 60 µl of PCR2 product from both 16S rDNA and ITS DNA pools were run into a 1.5% agarose gel. Then, the corresponding 16S rDNA and ITS DNA bands on the gel were carefully excised and processed to be purified using the manufacture's protocol of the Isolated II PCR and Gel Kit (Bioline®) illustrated in the Figure 3-15.

n) Quality control and molarity quantification

The final quality control of the pooled DNA libraries was to estimate and to ensure their molar concentrations according with the requirements demanded for being sequenced by the IIlumina platform which is 20 µl of a 10nM DNA solution. To ensure accurate results these measurements were made using a DNA 1000 protocol Assay on a Bioanalyzer 2100 instrument from Agilent

Technologies, INC. 2000[xxix]. The final pooled samples which molarity values exceeded 10nM DNA were diluted using the elution buffer from the purification kit. A set of three different dilutions with replicates were analysed by the Bioanalyzer to ensure molarity measurements.

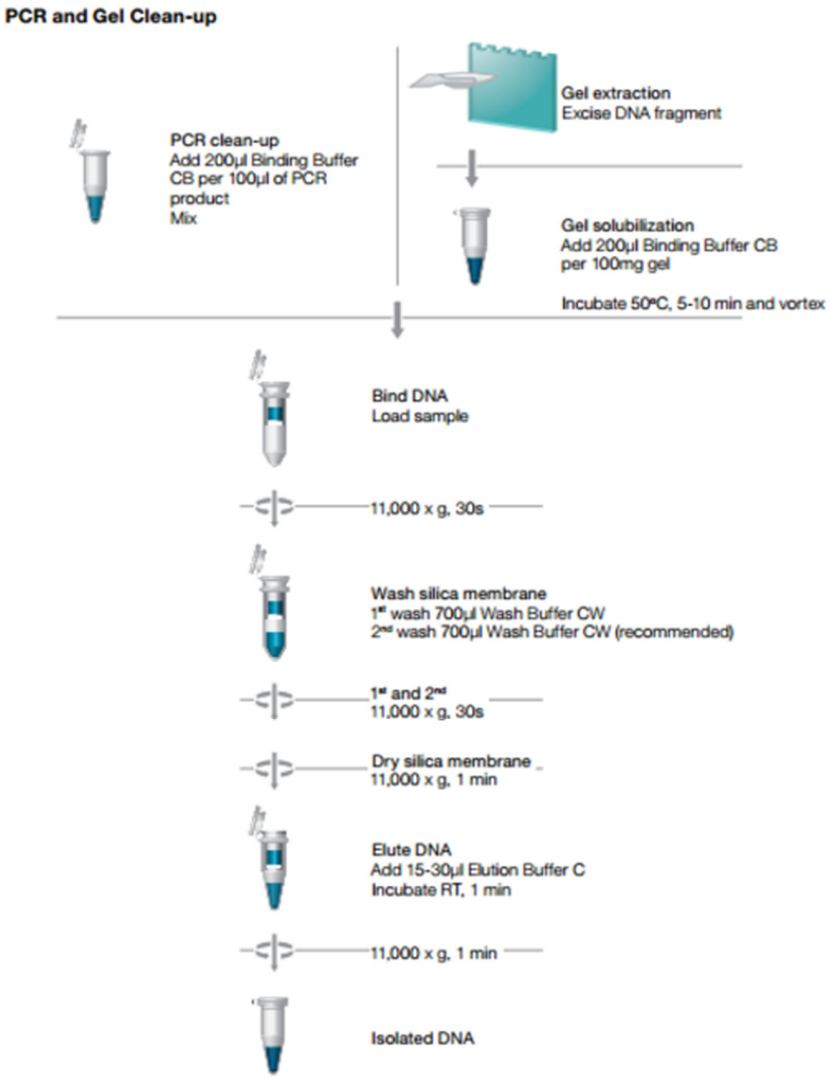


Figure 3-15. PCR clean-up based on gel extraction protocol for DNA pools amplicons purification.

[xxix] http://www.genomics.agilent.com/article.jsp?crumbAction=push&pageId=1628

By normalising the pooled samples at 10nM DNA, then 16S rDNA and ITS pools concentrations ended between 4 and 3 ng/µl, respectively (Table 3-8).

Table 3−8. DNA molarity and concentration conversion reference.

DNA size (bp)	Molarity	Concentration
600	10nM	4 ng/µl
500	10nM	3 ng/µl
400	10nM	2.5 ng/µl

o) 16S rDNA and ITS libraries mix

Finally, we prepared one DNA pool per each transect, for which, both 16S rDNA and ITS libraries were mixed up into one single 1.5 ml tube using aliquots of about 20 µl each. According with the concentrations values showed in Table 3-8 above, $NS_{\text{-transect}}$ pool and $WE_{\text{-transect}}$ pool ended concentrated at ~140 ng of microbial rDNA each.

These two libraries, $NS_{\text{-transect}}$ pool and $WE_{\text{-transect}}$ pool were sent to be sequenced by the Illumina platform placed at Monash University in Victoria, Australia. The quality of the DNA products was also tested in this laboratory prior to the sequence processing analyses.

Output from step 2…

Step 2 finished with a set of two microbial DNA pools contained into 1.5 ml tubes (one per each transect) which were sent to be sequenced on the Illumina platform.

STEP 3. AMPLICON ILLUMINA SEQUENCING

The soil microbial DNA samples were sequenced by Micromon Laboratory at Monash University in Victoria, Australia. The sequencing strategy used was a paired-end sequencing 2 X 300 bp on the Illumina MiSeq sequencer platform. As explained in the previous step, the microbial DNA for each transect was sequenced in two different pools: $NS_{\text{-transect}}$ pool *(*composed by 162 soil samples, blanks, and positive controls*)*, and $WE_{\text{-transect}}$ pool *(*composed by 132 soil samples, blanks, and positive controls*)*. Then, four fastq files were generated from the sequencing process per each transect: *forward and reverse target sequencing reads* and *forward and reverse index sequencing reads*.

A quality control and basic statistical analyses of the resulting raw sequencing reads was made using FastQC tool (V0.11.3) from the Babraham Bioinformatics Institute[xxx] (Andrews, 2010). This bioinformatics application provided a modular set of analyses to control the data derived from the high-throughput sequencing platform and report any misbehaviour in the sequencing calls. On the basis of the FastQC results, the raw sequence data were pre-treated by filtering the low quality data, *e.g.* sequences longer than ~210 and ~260 bp (Figure 3-16 and Figure 3-17).

The information enclosed in the Illumina files is briefly described in the Table 3-9, constructed with the reported FastQC analyses for both target and index (barcodes) readings. From this table we can be infer that the total number of sequences obtained from both transects sum up to

[xxx] http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/

47,115,100 sequences (approximately 31.6 GB of data).

Additionally, this table shows other statistical parameters which give a quick reference check on the quality of these sequence readings. For example, number of poor quality sequences, which were not found in these dataset; or percentage of guanine-cytosine (%GC), which being low or high can affect the accuracy of amplicon reads for microbial identification (Chen et al., 2013) but in this case results were within quite optimal ranges (*e.g.* 52 to 53% for target sequences) in both transects.

Table 3-9. Basics statistics on the raw data sequencing (.fastq) generated by MySeq platform from both DNA pool transects.

	NS-transect		WE-transect	
Measure	Forward Reads	Reverse Reads	Forward Reads	Reverse Reads
	Target sequences			
Total sequences	10,378,808	10,378,808	13,178,742	13,178,742
Sequences flagged as poor quality	0	0	0	0
Sequence length	35-301	35-301	35-301	35-301
%GC	52	53	52	53
	Index sequences (barcodes)			
Total sequences	10,378,808	10,378,808	13,178,742	13,178,742
Sequences flagged as poor quality	0	0	0	0
Sequence length	8	8	8	8
%GC	46	49	48	48
File Type: Conventional base calls				
Encoding: Sanger / Illumina 1.9				

To exemplify the quality control performed, refer to Figure 3-16 and Figure 3-17, which illustrate the results from one of several quality test modules analysed: *per base sequencing quality analyses*. Other module analyses included: per tile sequence quality, per sequence quality score, per sequence GC content distribution, sequence duplication levels, overrepresented sequences and *kmer* profiling, among other analyses.

Hence, the 'sequencing quality' plots shows good quality (green area), reasonable good quality (orange area) and poor quality (red area) of the base calls. Such report showed how the quality of the sequences decay for forward and reverse reads, respectively, after ~260 bp and ~ 210 bp lengths in both transects. This is because on most sequencing platforms, the quality of the reading is degraded as the run progresses, for which, it is common to see base calls falling into the orange area towards the end of the reading (Andrews, 2010).

All the quality information reported by these analyses was useful for evaluating the confidence on the datasets generated by the sequence processing as well as for taking decisions in downstream analyses when pre-treating the data and before proceeding with the analysis of these sequences for microbial identification.

Output from step 3…

Step 3 finished with a set of four .fastq files containing the raw sequencing data which were analysed for microbial identification.

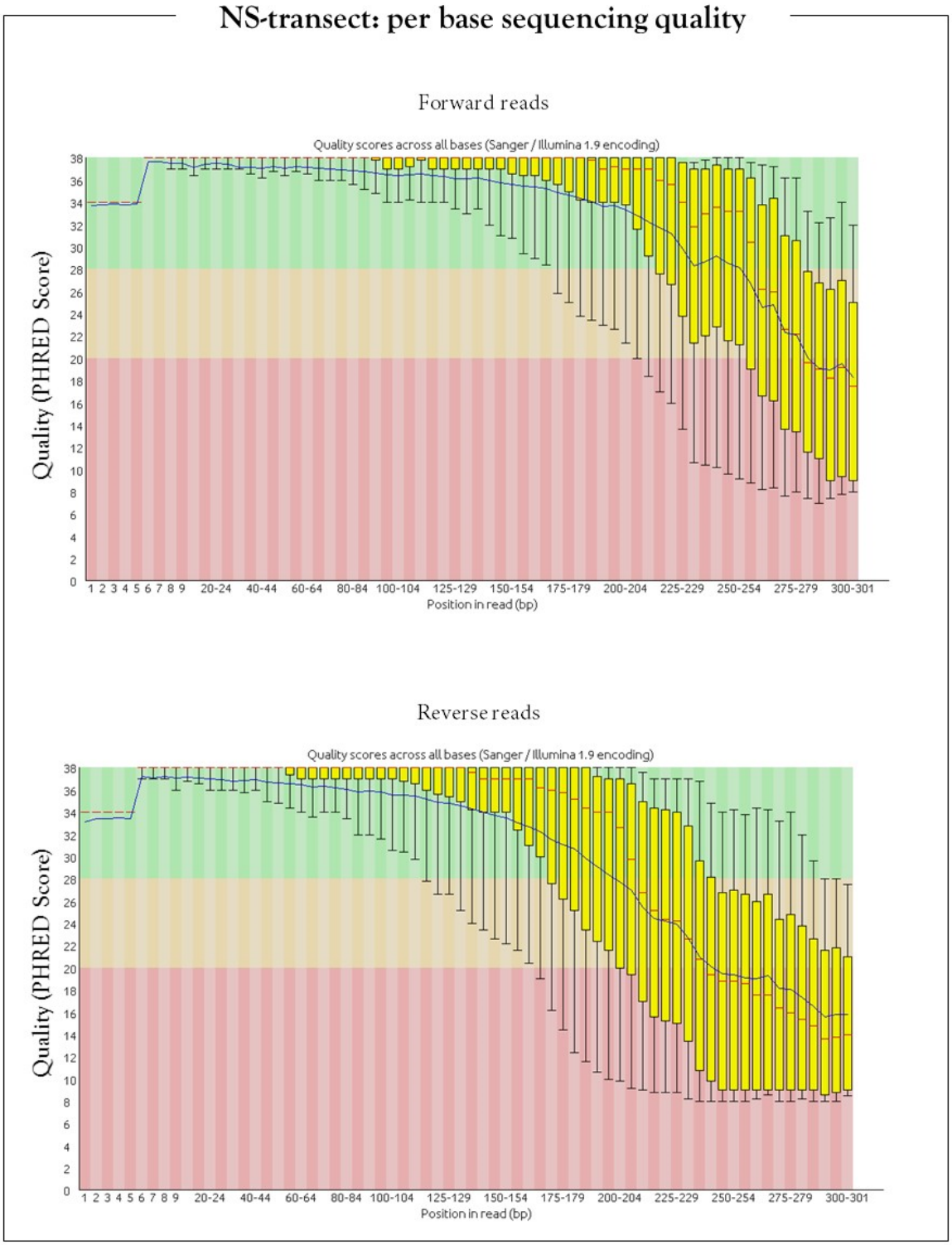


Figure 3-16. Quality scores across forward and reverse reads reported by FastQC test for NS-transect.

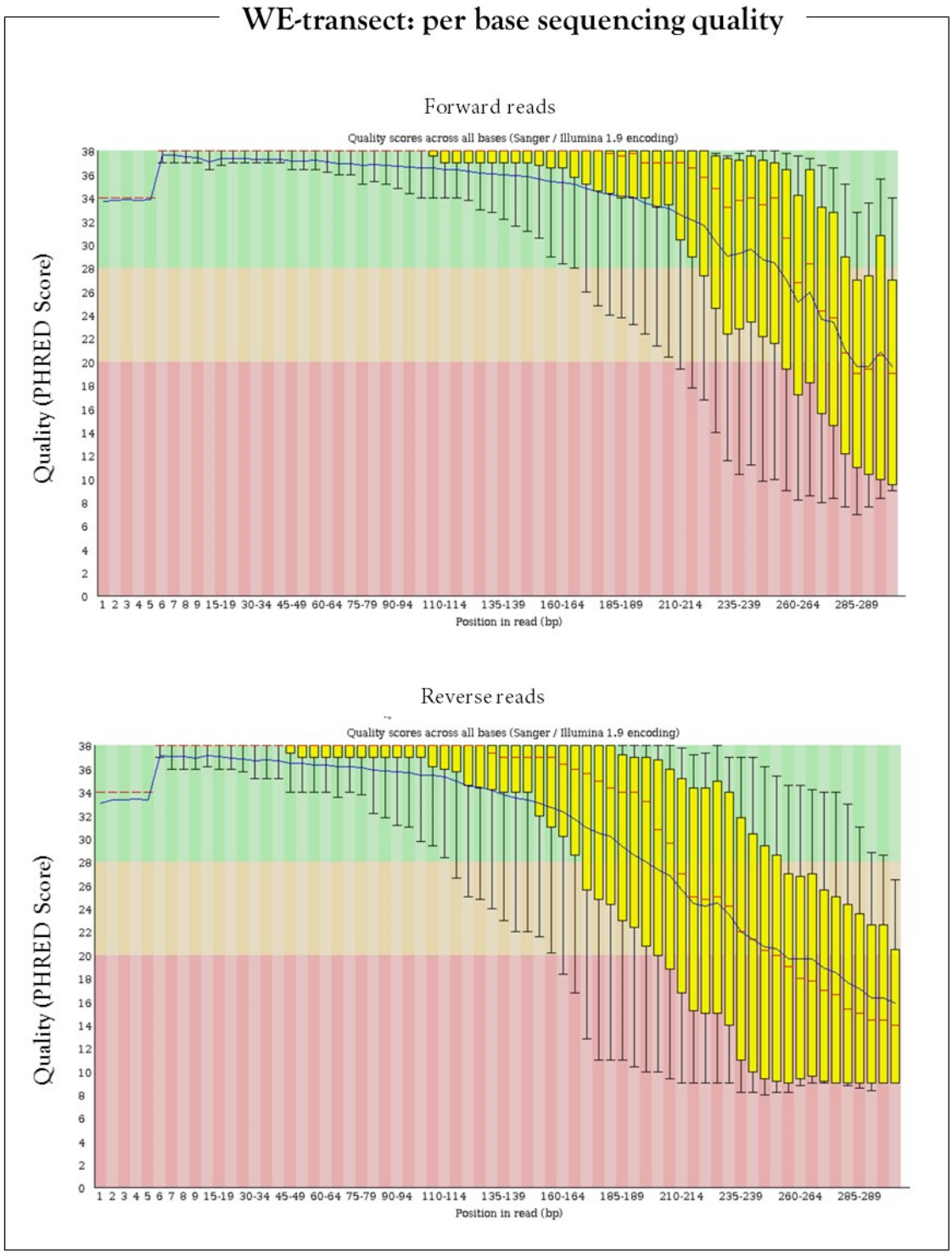


Figure 3-17. Quality scores across forward and reverse reads reported by FastQC test for *WS*-transect.

STEP 4. MICROBIAL IDENTITY-QUANTIFICATION

Soil microorganisms were identified and quantified by processing the Illumina DNA sequencing reads (.fastq files) through a set of open-source bioinformatics applications that allowed the identification of OTUs. Once qualified the raw sequencing dataset, forwards and reverse raw sequencing reads files were filtered prior construction of the OTUs database. This data processing was applied differently for 16S rDNA and ITS libraries in accordance with bioinformatics pipelines particularly developed for analysing each taxonomic group, *i.e. bacteria/archaea* and *fungi*.

Therefore, the final OTUs database of *archaea* and *bacteria* (from 16S rDNA seqs) were obtained using open-source pipelines available in QIIME (Quantitative Insights Into Microbial Ecology, *version 1.9.1*) introduced by Caporaso et al., 2010. On its own, the *fungi* OTUs database (from ITS seqs) was constructed by combining the QIIME and PIPITS pipeline, which is also an open and automated bioinformatics application specifically designed for Illumina ITS sequencing analyses (Gweon et al., 2015). In both cases, the paired-end readings, *i.e.* forward and reverse sequences, were merged by means of PEAR software, a read merger tool specially developed for raw Illumina paired-end from target amplicons of variable lengths (Zhang et al., 2014).

Despite this differentiation in the data processing for 16S and ITS identification, the entire procedure to obtain the OTUs database for identity-abundance of the soil microorganisms involved is described in Figure 3-18.

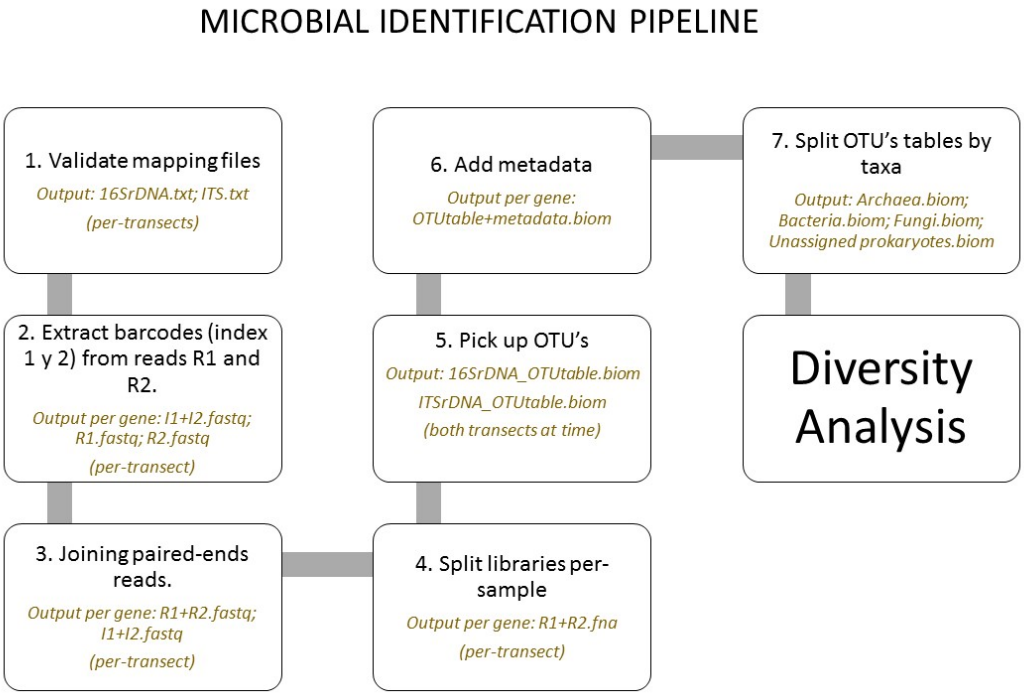


Figure 3–18. Bioinformatics workflow describing the sequencing reads analyses for microbial identification.

The bioinformatics procedures and the specific configurations applied in each of these modules are specified below:

From Illumina paired-end reads sequencing library to microbial OTU tables:

1. *Creating and validating mapping files:* 16SrDNA and ITS mapping files (.txt) were created to provide per-sample metadata by which associate samples ID's to their corresponding barcodes (index).

2. *Extracting index barcode:* the forward and reverse barcodes of 8 bp indexing the DNA samples were extracted from the target sequences and merged into other fastq files.

3. *Joining paired-ends reads*: forwards and reverse sequencing reads (.fastq) were merged into one single read to increase the overall read length. This merging method was worked using a cross-platform processing between QIIME and PEAR software.

4. *Splitting sequencing libraries per sample and filtering quality*: $NS_{-transect}$ and $WE_{-transect}$ sequencing libraries (.fastq) contained multiplexed DNA sequences that must be de-multiplexed by distributing the corresponding sequences to each soil sample, generating new files (.fna) per- taxa and transect. This procedure, which was run in QIIME for both 16S and ITS sequences libraries, involved the primary quality filter on the sequencing reads. The quality-filtering strategy was applied according with the optimal quality calibration recommended by Bokulich et al., (2013), *i.e.* with thresholds of $r=3$; $q=3$; $p=0.75$ and $n=0$, where q corresponded with the lowest quality score, p is the minimum number of consecutive high-quality base calls to retain a read (as % of total read length) and n is the maximum number of ambiguous characters allowed in a sequence. These parameters allowed to extract high-quality data to facilitate its interpretation even though, there was a secondary quality-filtering run after picking up the final OTUs.

5. *Picking up OTUs:* microbial identification was processed by clustering the target sequence reads (.fna) into OTUs (.biom) using a reference collection. OTUs from 16S rDNA sequences were picked up using an open-reference strategy in QIIME. In this pipeline, the taxonomic assignment has been made against the Greengenes reference

dataset (release 13_8) (DeSantis et al., 2006) using UCLUST algorithm as the clustering method at a 97% sequence similarity threshold, *i.e.* species level (DeSantis et al., 2006; Edgar, 2010). Then, the sequences were aligned via the PYNAST (version 1.2.2) algorithm method to infer phylogeny and region of similarities across the sequences (Caporaso et al., 2010a). By this strategy, unassigned sequences were classified as *novo*.

In the f*ungi* identification, the OTUs from ITS sequences were picked up following a PIPIT pipeline. By means of this pipeline, the taxonomy was assigned against the UNITE fungal ITS reference dataset (Xu et al., 2015) with the RDP classifier (version 2.9) (Wang et al., 2007) for clustering at a 97% similarity threshold. In both cases, during this taxonomic assignment, $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$ were run together in order to use the same representative sequences for the taxonomic identification of the OTUs.

6. *Adding metadata to the sequencing library:* additional information related to sampling sites and other observations were added to the OTUs tables.

7. *Splitting OTUs tables by taxa:* the 16SrDNA sequencing OTU table (16Sotu_table.biom) was separated into *archaea, bacteria* and *unassigned prokaryotes* files tables for downstream diversity analyses. The unassigned prokaryotes have been considered in downstream diversity analyses since this is a quite numerous group. This group of microbes still lack taxonomic classification but in the near future could be part of the recently proposed candidate phyla Radiation (Hug et al., 2016). However, this

information will be defined in future investigations, and within the context of this research they will be referred as *unassigned prokaryotes*.

Finally, soil microorganisms were count/identified and biological observation matrix (.biom) tables was created for each taxa, *i.e. fungi bacteria*, *archaea*, as well as for those unassigned *prokaryotes*. These biom tables are described in downstream analyses when analysing microbial alpha and beta diversity in the Chapters 4 and 5.

Output from step 4…

Step 4 concluded with a set of four .biom tables (archaea, bacteria, fungi, unassigned prokaryotes) containing the microbial identity/abundance per each sampling site composing our study area.

SOIL PHYSICOCHEMICAL CHARACTERISATION

The soil physiochemical characterization has been made using both laboratory and pedometrics techniques to provide additional pedodiversity information for further analyses. As specified in Chapter 2, a set of 196 soil samples were taken for the soil chemico-physical characterisation. This samples came from 0-5 and 5-10 cm depth for each of the 49 sampling sites composing $NS_{-transect}$ (108 soil samples) and $WE_{-transect}$ (88 soil samples). In addition to these soil samples, an extra set of 98 cores (100 cm depth) were taken to determine the soil morphological horizonation of soil profiles. On the basis of these set of soil samples, the soil physiochemical analyses were made using conventional (i) chemical and (ii) physical laboratory measurements (Step 1) and other pedometrics approaches (Step 2) including the use of Vis-NIR/mid-IR spectral absorbance for (iii) the prediction of certain soil physical and chemical properties, (iv) the assessing of soil aggregates stability by image recognition and, (v) estimation of the soil spectral derived horizonation.

The number of replicates used for these measurements were variable per property analysed and the different techniques/approaches employed (e.g. laboratory measures, Vis-NIR/mid-IR spectral absorbance, etc.). This information and other details are given below when describing each of the procedures applied. The soil physiochemical properties results are shown below (Step3).

STEP 1. LABORATORY MEASUREMENTS

(i) Soil chemical properties

The soil chemical properties characterisation and methods used for this purpose were: Ammonium-N (NH_4-N) and Nitrate-N (NO_3-N) in mg kg^{-1}, Exchangeable Cations (Ca, Mg, Na, K, Al) in cmol $^{(+)}$ kg^{-1} and Effective Cation Exchange Capacity (ECEC) in cmol$^{(+)}$ kg^{-1} applying Rayment and Higginson (1992) protocols; Total Carbon content (TC) and Nitrogen content (%N) by dry combustion using LECO® instrument manufacture's procedure; Extractable Phosphorus (P) and Potassium (K) in mg kg^{-1} by Colwell (1963) method; and Electric conductivity (EC) in dS/m and pH in water. All the laboratory routines were performed by CSBP Laboratory Ltd., in Western Australia.

(ii) Soil physical properties

The same set of sample was used to perform physical analyses of particle size (clay, silt and sand) in percentage mass by the hydrometer method (Bouyoucos, 1962). These measurements were estimated in the Soil Security Laboratory at The University of Sydney, Australia.

STEP 2. OTHER PEDOMETRICS MEASUREMENTS

(iii) Soil Vis-NIR/mid-IR spectral absorbance measurements

Soil spectral curves were constructed using mid-infrared spectroscopy (mid-IR) and visible and near infrared spectroscopy (Vis-NIR) ranging between wavenumbers 600 – 4,000 cm^{-1} to 4,000

– 25,000 cm^{-1}, respectively. The mid-IR absorbance was measured using an absorbance FT-IR Spectrometer TENSOR 37 with a HTS-XT Microplate Reader Bruker®. In the case of Vis-NIR absorbance measurements, they were obtained by an ASP 350-2500 AgriSpec with a Spectralon® panel as absolute white. Vis-NIR and mid-IR were estimated for all the 196 soil samples and five replicates measurement were taken per each sample.

(iv) Aggregate stability by soil slaking index metric

The aggregation is qualitatively related to microbial community structure and activity. Aggregate protect soil organic matter, limit oxygen diffusion, regulate water flow, control nutrient adsorption/desorption and also diminish the effect of run-off and erosion (Six et al., 2004). Slaking index is a physical assessment which evaluates the soil aggregate instability. In general terms, the protocol works based on an image recognition technique. This is a novel technique that was developed within the context of the study area involved in this research, and it has been recently introduced by Fajardo et al., 2016. Thus, major specifications related to the aggregate measuments can be consulted directly in this article.

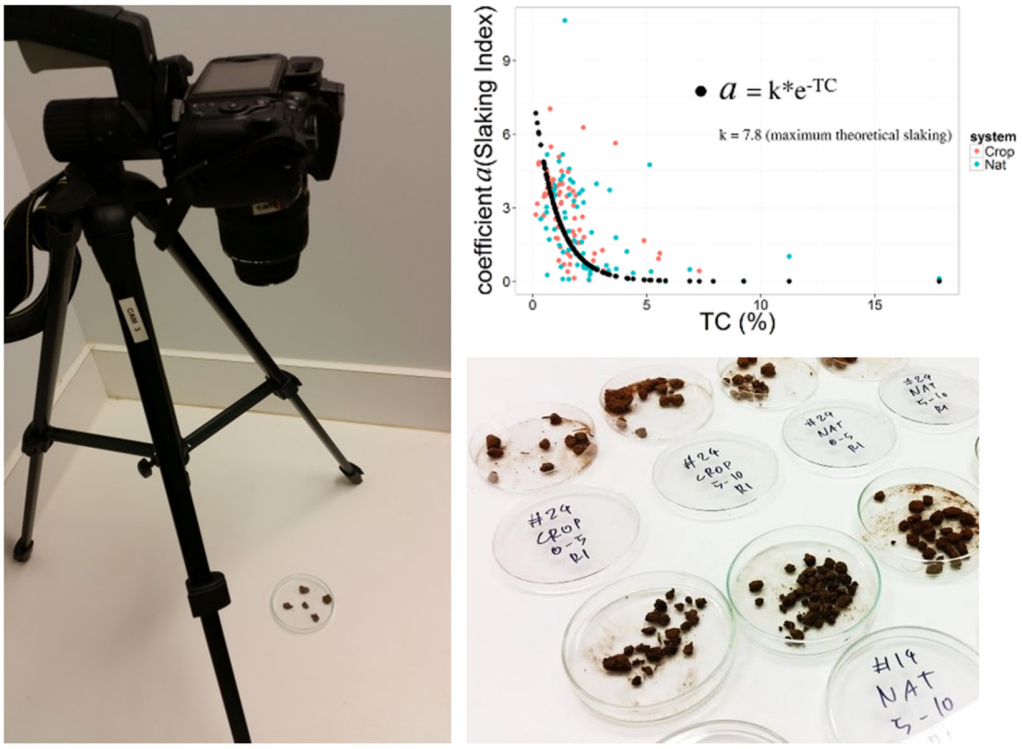


Figure 3−19. Soil slaking assessed by image recognition. *Photo*: Courtesy Of Fajardo et al., 2016. The graph in the image exposes the soil disaggregation dynamics in water.

The slaking measurements were directly measured from a set of 148 out of the original 196 soil samples, since certain samples lacked stable aggregates (*e.g.* sandy soil) or their aggregates were excessively hydrophobic due to high organic matter content (Fajardo et al., 2016b). The final slaking index values (*i.e.* coefficient 'a') were calculated from an average of five aggregates per sample. As high value represents more unstable aggregates and, consequently, weaker the soil structure. The remaining 48 soil samples were estimated from the mid-IR spectral dataset (R^2 0.6; RMSE 1.04 SI and BIAS 0.05) (Figure 3-19).

(v) Spectrally derived soil horizons

A total of 98 soil cores extracted from 100 cm depth from each of the 49 sampling site were analysed by Vis-NIR spectroscopy to generate soil spectral profiles each 2 cm. Then, the spectral data were analysed to obtain two morphological descriptors of the soil horizonation features in the profile: (a) *number of horizons (N-Hor) along the profile* and (b) *thickness of the first horizon* (*T-Hor*) (Figure 3-20). Such morphological descriptors were estimated using a Vis-NIR fuzzy clustering approach introduced and described in Fajardo et al., in 2014. Both number of horizons and thickness of the first horizon are additional information to assess pedodiversity in terms of variation of the soil composition in depth.

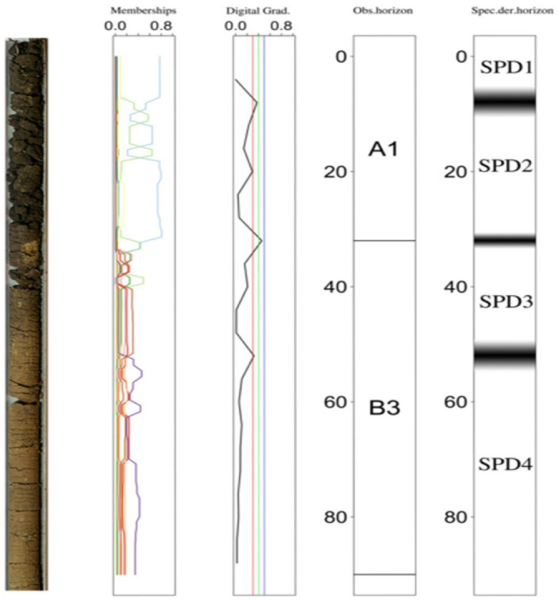


Figure 3-20. Vis - NIR fuzzy clustering method apply to recognise soil horizons. *Photo*: Courtesy Of Fajardo et al., 2014, which shows the digital gradient of a soil core and the boundaries distinctness derived and related to the horizon thickness boundary.

STEP 3. SOIL PHYSICOCHEMICAL RESULTS

The range of soil physicochemical properties values are shown below for each transects and ecosystem. Table 3-10 and Table 3-11 displays the range of soil properties measured in sampling sites along $NS_{-transect}$ and $WE_{-transect}$, respectively.

Table 3−10. Soil physicochemical properties ranges found in $NS_{-transect}$.

Property	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Min	1st Quartile	Median	Mean	3rd Quartile	Max
	Natural Ecosystems						Cropping Ecosystems					
NH_4-N (mg Kg^{-1})	1.00	4.00	6.00	9.07	9.00	43.00	1.00	3.00	4.00	5.15	7.00	17.00
NO_3-N (mg Kg^{-1})	1.00	2.00	6.00	12.72	10.00	123.00	1.00	6.00	13.50	20.57	23.75	110.00
P (mg Kg^{-1})	4.00	8.00	18.00	26.63	40.25	91.00	6.00	23.50	37.00	39.89	49.75	99.00
K (mg Kg^{-1})	99.00	258.20	393.50	384.10	503.80	720.00	74.00	260.50	359.50	385.20	500.00	814.00
EC (dS m^{-1})	0.01	0.04	0.06	0.08	0.09	0.24	0.02	0.04	0.07	0.08	0.11	0.22
pH ($CaCl_2$)	3.80	4.50	5.00	5.25	5.90	7.50	4.40	4.90	5.50	5.60	6.08	7.60
pH (H_2O)	4.70	5.30	6.00	6.07	6.60	8.40	5.10	5.80	6.20	6.39	6.70	8.60
Exc. Al (cmol (+) kg^{-1})	0.01	0.06	0.10	0.31	0.41	1.79	0.01	0.04	0.05	0.12	0.11	1.24
Exc. Ca (cmol (+) kg^{-1})	0.27	2.30	5.26	8.04	10.76	30.64	0.62	4.12	6.25	8.58	10.47	24.05
Exc. Mg (cmol (+) kg^{-1})	0.29	1.04	2.32	3.37	5.04	13.12	0.27	0.81	1.75	2.91	2.94	12.63
Exc. K (cmol (+) kg^{-1})	0.25	0.62	0.99	0.96	1.24	1.84	0.19	0.65	0.88	0.97	1.29	2.06
Exc. Na (cmol (+) kg^{-1})	0.01	0.02	0.09	0.24	0.34	1.68	0.01	0.04	0.13	0.24	0.33	1.26
ECEC (cmol (+) kg^{-1})	1.70	5.10	8.95	12.92	17.38	38.30	2.70	6.90	8.80	12.82	13.87	36.60
TC (%)	0.61	1.47	2.20	2.81	2.71	17.83	0.48	1.19	1.59	1.76	2.06	7.31
TN (%)	0.07	0.11	0.16	0.21	0.22	1.21	0.04	0.10	0.14	0.16	0.19	0.57
Clay (%)	3.05	12.94	18.31	23.23	25.69	64.00	6.53	13.23	16.40	22.80	23.20	65.11
SI	0.00	0.55	1.45	1.68	2.75	5.17	0.05	1.33	2.06	2.44	3.54	6.27
N-Hor	1	4	5	5	6	9	1	3	4	4	5	7
T-Hor (cm)	1.00	3.00	5.00	6.22	8.75	15.00	1.00	2.25	4.00	5.41	7.50	15.00

These soil physical and chemical data will be analysed and discussed in subsequent chapters together with alpha and beta diversity analyses and in the context of the soil gradients characterising both the $NS_{-transect}$ and $WE_{-transect}$.

Table 3−11. Soil physicochemical properties ranges found in $WE_{-transect}$.

Property	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Min	1st Quartile	Median	Mean	3rd Quartile	Max
	Natural Ecosystems						Cropping Ecosystems					
NH_4-N (mg Kg^{-1})	0.1	1.0	4.0	16.4	9.0	195.0	0.1	1.0	3.0	11.4	13.5	76.0
NO_3-N (mg Kg^{-1})	1.0	5.0	13.0	28.5	38.3	202.0	1.0	5.8	17.5	37.8	45.3	354.0
P (mg Kg^{-1})	3.0	17.5	30.5	52.4	59.5	424.0	3.0	16.0	26.0	52.1	48.5	273.0
K (mg Kg^{-1})	35.0	358.0	449.5	559.0	787.2	1377.0	35.0	281.2	476.0	487.8	614.2	962.0
EC (dS m^{-1})	0.0	0.0	0.1	0.1	0.1	0.5	0.0	0.1	0.1	0.1	0.1	0.5
pH ($CaCl_2$)	4.2	5.2	6.0	6.0	6.9	7.6	4.2	5.0	6.1	6.1	7.0	7.8
pH (H_2O)	5.0	5.8	6.6	6.7	7.6	8.9	5.0	5.7	6.9	6.8	8.0	8.6
Exc. Al (cmol (+) kg^{-1})	0.0	0.1	0.1	0.2	0.1	4.0	0.0	0.1	0.1	0.1	0.1	0.4
Exc. Ca (cmol (+) kg^{-1})	0.9	4.8	11.4	12.6	20.8	27.2	0.9	4.5	13.5	11.6	17.4	22.2
Exc. Mg (cmol (+) kg^{-1})	0.3	1.5	3.8	4.8	7.1	14.7	0.3	1.4	4.4	4.7	7.2	13.8
Exc. K (cmol (+) kg^{-1})	0.1	0.7	1.1	1.3	1.7	3.0	0.1	0.6	1.2	1.1	1.6	2.1
Exc. Na (cmol (+) kg^{-1})	0.0	0.0	0.1	0.4	0.4	2.6	0.0	0.0	0.2	0.5	0.7	3.9
ECEC (cmol (+) kg^{-1})	1.9	8.1	16.1	19.3	30.6	43.0	1.9	6.6	18.8	18.1	28.7	35.6
TC (%)	0.4	0.9	1.7	4.6	5.6	31.2	0.4	0.6	1.1	2.1	3.2	10.5
TN (%)	0.0	0.1	0.2	0.3	0.4	2.0	0.0	0.1	0.1	0.2	0.3	0.8
Clay (%)	3.2	13.3	29.2	28.3	40.2	53.6	3.2	24.9	34.8	32.6	42.5	51.7
SI	0.0	1.7	3.2	2.9	4.0	10.6	0.0	1.9	3.3	3.1	4.1	7.0
N-Hor	1	3	4	4	4	8	1	2	3	3	4	7
T-Hor (cm)	1.0	3.0	5.0	7.0	12.0	15.0	1.0	4.0	6.0	7.3	11.0	15.0

In the next chapter…

In the next chapter 4, soil microbial alpha diversity is estimated and afterward mapped according its relation with soil and other environmental gradients along both transects.

REFERENCES

Aaij, C., Borst, P., 1972. The gel electrophoresis of DNA. BBA Sect. Nucleic Acids Protein Synth. 269, 192–200. doi:10.1016/0005-2787(72)90426-1

Andrews, S., 2010. FASTQC A Quality Control tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360, 1935–1943. doi:10.1098/rstb.2005.1725

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods 10, 57–59. doi:10.1038/nmeth.2276

Bouyoucos, G.J., 1962. Hydrometer Method Improved for Making Particle Size Analyses of Soil. Agron. J. 54, 464. doi:10.2134/agronj1962.00021962005400050028x

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinforma. 26, 266–267. doi:10.1093/bioinformatics/btp636

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G. a, Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C. a, Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W. a, Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R.,

2010b. Correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat. Publ. Gr. 7, 335–336. doi:10.1038/nmeth0510-335

Chen, Y.C., Liu, T., Yu, C.H., Chiang, T.Y., Hwang, C.C., 2013. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. PLoS One 8. doi:10.1371/journal.pone.0062856

Colwell, J.D., 1963. The estimation of the phosphorus fertilizer requirements of wheat in southern New South Wales by soil analysis. Aust. Jour Exptl Agric Anim. Husb 3, 190–197. doi:10.1071/ea9630190

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–72. doi:10.1128/AEM.03006-05

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinforma. 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., Ravel, J., 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2, 6. doi:10.1186/2049-2618-2-6

Fajardo, M., McBratney, A., Whelan, B., 2016a. Fuzzy clustering of Vis-NIR spectra for the objective recognition of soil morphological horizons in soil profiles. Geoderma 263, 244–253. doi:10.1016/j.geoderma.2015.05.010

Fajardo, M., McBratney, A.B., Field, D.J., Minasny, B., 2016b. Soil slaking assessment using image recognition. Soil Tillage Res. 163, 119–129. doi:10.1016/j.still.2016.05.018

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J. a, Wall, D.H., Caporaso, J.G., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. U. S. A. 109, 21390–5. doi:10.1073/pnas.1215210110

Gweon, H.S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D.S., Griffiths, R.I., Schonrogge, K., 2015. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Methods Ecol. Evol. 6, 973–980. doi:10.1111/2041-210X.12399

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Kotaro, I., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree and life's diversity. Nat. Microbiol. 1, 16048. doi:10.1038/nmicrobiol.2016.48

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., Neufeld, J.D., 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Appl. Environ. Microbiol. 80, 5717–5722. doi:10.1128/AEM.01451-14

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analysing amplicon sequence data on the miseq illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120. doi:10.1128/AEM.01043-13

Larsen, P., Hamada, Y., Gilbert, J., 2012. Modeling microbial communities: Current, developing, and future technologies for predicting microbial community interaction. J. Biotechnol. 160, 17–24. doi:10.1016/j.jbiotec.2012.03.009

Levy-Booth, D.J., Campbell, R.G., Gulden, R.H., Hart, M.M., Powell, J.R., Klironomos, J.N., Pauls,

K.P., Swanton, C.J., Trevors, J.T., Dunfield, K.E., 2007. Cycling of extracellular DNA in the soil environment. Soil Biol. Biochem. 39, 2977–2991. doi:10.1016/j.soilbio.2007.06.020

Mendoza, M.L.Z., Sicheritz-Pont??n, T., Thomas Gilbert, M.P., 2014. Environmental genes and genomes: Understanding the differences and challenges in the approaches and software for their analyses. Brief. Bioinform. 16, 745–758. doi:10.1093/bib/bbv001

MO BIO Laboratories, 2014. PowerSoil® DNA Isolation Kit: Instruction Manual 24.

Nesme, J., Achouak, W., Agathos, S.N., Bailey, M., Baldrian, P., Heulin, T., Jansson, J.K., Jurkevitch, E., Kruus, K., Kowalchuk, G.A., Lagares, A., Lappin-scott, H., Lemanceau, P., Le, D., Mandic-mulec, I., Murrell, J.C., Myrold, D.D., Nalin, R., Nannipieri, P., Neufeld, J.D., Parnell, J.J., Pylro, V., Ramos, J.L., Roesch, F.W., Schloter, M., Schleper, C., Sczyrba, A., Sessitsch, A., Tebbe, C., Topp, E., Tsiamis, G., Elsas, J.D. Van, Keulen, G. Van, Widmer, F., Wagner, M., Zhang, T., Zhang, X., Zhao, L., Zhu, Y., Vogel, T.M., Simonet, P., Genomics, E.M., Lyon, E.C. De, Environment, N., Road, M., Science, F., Directorate, C.S., Biology, E., Plata, L., Plata, L., Science, E., Science, B., Ecological, N., Network, O., Gabriel, S., 2016. Back to the future of soil metagenomics. Front. Microbiol. doi:10.3389/fmicb.2016.00073

Rayment, G.E., Higginson, F.R., 1992. Australian Laboratory Handbook of Soil and Water Chemical Methods. Inkata Press.

Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: Evaluating the illumina platform. PLoS One 10, 1–18. doi:10.1371/journal.pone.0116955

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31. doi:10.1016/j.still.2004.03.008

Southern, E.M., 1979. Measurement of DNA length by gel electrophoresis. Anal. Biochem. 100, 319–323. doi:10.1016/0003-2697(79)90235-5

Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. Mol. Ecol. 21, 1789–1793. doi:10.1111/j.1365-294X.2012.05542.x

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., Nishijima, M., Miseq, I., 2014. PLOS ONE Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLoS One 9, e105592. doi:10.1371/journal.pone.0105592

Toju, H., Tanabe, A.S., Yamamoto, S., Sato, H., 2012. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS One 7. doi:10.1371/journal.pone.0040863

Torsvik, V., Daae, F.L., Sandaa, R.A., Øvreås, L., 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. J. Biotechnol. 64, 53–62. doi:10.1016/S0168-1656(98)00103-5

Torsvik L., V.O., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5, 240–245. doi:http://dx.doi.org/10.1016/S1369-5274(02)00324-7

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–7. doi:10.1128/AEM.00062-07

Xu, X., Passey, T., Wei, F., Saville, R., Harrison, R.J., 2015. Amplicon-based metagenomics identified candidate organisms in soils that caused yield decline in strawberry. Hortic. Res. 2, 15022. doi:10.1038/hortres.2015.22

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B.,

Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645. doi:10.1038/nrmicro3330

Zarraonaindia, I., Smith, D.P., Gilbert, J. a., 2013. Beyond the genome: Community-level analysis of the microbial world. Biol. Philos. 28, 261–282. doi:10.1007/s10539-012-9357-8

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620. doi:10.1093/bioinformatics/btt593

Chapter 4

Soil microbial *α-diversity* across New South Wales

Outlines

Microbial α-diversity has been assessed at each sampling site along the $NS_{-transect}$ and $WE_{-transect}$. These analyses and its corresponding results are presented on the basis of microbial relative abundance distributions, taxonomic compositions and phylogenetic diversity. To understand changes in the microbial diversity patterns across environmental gradients, microbial α-diversity has been evaluated in association with soil physicochemical and other environmental attributes. The results are presented in this chapter together with microbial models and predictions based on them. We demonstrate the close relationship between soil microbial communities and multiple soil attributes.

Summary

Soil microbial diversity and soil physicochemical attribute relationships were assesed across different agroecological zones in NSW. In this context, soil *bacteria*, *archaea* and *fungi* α-diversity were evaluated across the 49 nine sampling sites from the $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$. Both microbial and soil parameters were analysed in paired disturbed/managed and natural/undisturbed ecosystems. Finally, the complex associations found between biotic-abiotic factors were integrated by modelling and mapping of microbial diversity across all of NSW.

The microbial DNA sequencing dataset on which diversity and taxonomic composition were assigned as well as procedures followed during these estimations, are described prior to analyse α-diversity. In this respect, the extent of the microbial dataset used for diversity calculations at 97% similarity resolution comprised 11,557,499 sequences (filtered) and 423,740 OTUs. Taxonomic and diversity measurements of richness (*e.g.* Chao1, Observed OTUs and PD) and evenness (*e.g.* H', 1-D) as well as the distribution of the relative abundances (rarity, commonness) were analysed upon the sequencing dataset. The estimations and resulting microbial taxonomic composition, relative abundances and diversity richness/evenness data have been then evaluated in the context of our study area, *i.e.* distributed across our 49 sampling sites along the environmental gradients of both transects.

A first insight into *archaea*, *bacteria* and *fungi* abundance/diversity are displayed to scope their fluctuations in the study area. Coarsely, on the basis of the number of sequences, we found that fungal community was significantly more abundant than the bacterial community but the latter

were much more diverse. Archaeal community was the least abundant/diverse and the most fluctuating across the study area. Reasonably, due to methodological inconsistences (e.g. gene amplifications, DNA extraction efficiency) these estimation cannot be considered adequately accurate but it depicts a general view of the microbial populations. Taxonomic diversity is illustrated across the sampling sites and analysed in accordance with the effect of land use (disturbed and undisturbed). In this regard, the most abundant phyla we found were *Actinobacteria (bacteria)*, *Ascomycota (fungi)* and *Crenarchaeota (archaea)*. The relative abundances (*e.g.* commonness and rarity) and tendencies of diversity patterns of each kingdom were analysed along the latitudinal and longitudinal gradient in both disturbed and undisturbed ecosystems. We observed, for example, that contrasting *bacteria* and *fungi* pattern only archaeal diversity tended to increase towards warmer zones (Queensland border). Finally, the microbial information was analysed in relation with abiotic parameters in which we emphasised soil physicochemical properties. These analyses were organized from more simple relations (linear, quadratic) to more complex scenarios. On the basis of multivariate relations (e.g. PCA, bootstrap regression modelling and, mapping) found between microbial diversity and abiotic factors (soil properties and other covariates), α-diversity of *archaea*, *bacteria* and *fungi* were predicted and mapped for NSW.

Our results have highlighted the close synergy between microbial community structure and soil physiochemical attributes but foremost within the extent of soil heterogeneity, i.e. pedodiversity. We demonstrated that in the complexity of this biotic-abiotic relationship, there is a set of soil properties promoting microbial diversity patterns rather than single attributes.

MEASURING MICROBIAL DIVERSITY

DEFINING THE MICROBIAL DATASET

The DNA sequencing dataset on which microbial abundances and microbial diversity metrics were afterwards calculated are described below.

DNA sequencing dataset

Descriptive statistics

Soil microbial diversity was analysed in terms of *prokaryotes* (*bacteria*, *archaea* and other *unassigned*) and *eukaryotes* (*fungi*) from a quality-filtered DNA sequencing dataset of 12,104,155 sequences from all 294 soil surface samples (0-5 cm depth) as described in Chapter 3. In terms of the amount of DNA sequences, this dataset is distributed into 0.2% *archaea*, 2.5% *unassigned prokaryotes*, 40.6% *bacteria* and 56.7% *fungi*. On average, from each soil sample we obtained 78.3 sequences of *archaea*, 1,051.5 sequences of *unassigned prokaryotes*, 17,689 sequences of *bacteria* and 23,325.1 sequences of *fungi, i.e.* 42,144 sequences per soil sample (Table 4-1). In these databases we also found 423,740 phylotypes (OTUs) at the 97% similarity level. These phylotypes represent a general view as to the extent of the microbial richness found in our entire study area ($NS_{\text{-transect}}$ and $WE_{\text{-transect}}$). These phylotypes distributed per taxa comprise of 496 *OTUs of archaea (*found in 77 of 294 soil samples*)*, 76,896 OTUs of *unassigned prokaryotes* (from in 122 of 294 soil samples), 323,223 OTUs of *bacteria (*from in 228 of 294 soil samples*)* and 23,125

OTUs of *fungi* (found in 247 of 294 soil samples). In spite of the cut-off being at the minimum level of sampling effort (*e.g.* 10,000 seqs for *bacteria* and *fungi* as described in Chapter 4), the amount of these OTUs observations did not diminish. In this case, the range of sequences per sample within the whole dataset was from 21,332 to 151,721. Finally, the filtered dataset based on the number of observed OTUs was comprised by 0.1% of *archaea*; 18.1% of *unassigned prokaryotes*; 76.3% of *bacteria* and 5.5% of *fungi* – this was the dataset used for diversity analysis.

Table 4−1. Descriptive statistics for the taxa sequencing datasets obtained per soil sample before and after cutting−off at the minimum level of sampling effort.

Taxa	Nº of soil samples	Nº of OTUs	Nº of sequences	Min	Max	Median	Mean	Std. Dev.
			Sequencing dataset at 97% similarity					
Archaea	293	496	22,945	0	921	33	78.3	127.1
Unassig. Prok.	293	76,896	308,103	0	5,325	785	1,051.5	905.8
Bacteria	293	323,223	4,915,523	1	50,546	17,689	16,776.5	9,045.3
Fungi	294	23,125	6,857,584	3	94,929	21,141	23,325.1	14,504.0
Full community	294	423,740	12,104,155	–	–	–	–	–
			*Sequencing dataset at 97% similarity after the cut-off at the level of sampling effort**					
Archaea	77	496	17,583	80	921	158	228.4	172.3
Unassig. Prok.	122	76,896	232,105	1004	5,325	1,696	1,902.5	779.5
Bacteria	228	323,223	4,666,261	10,139	50,546	19,622	20,466.1	6,334.6
Fungi	247	23,125	6,641,550	10,109	94,929	23,555	26,888.9	12,970.4
Full community	247	423,740	11,557,499	–	–	–	–	–

*Depth of sampling cut−offs: *archaea*=80 seqs; unassigned prokaryotes=1,000 seqs; *bacteria* and *fungi*=10,000 seqs.

In light of these results, it can be seen that *fungi* is the most abundant but *bacteria* is the richest group in our dataset. By contrast, *archaea* happens to be the smallest group in this dataset, in terms of both richness and abundance. These results does not necessarily reflect the impact of environmental variables on our soils samples since the efficiencies on the genes amplifications and other methodological bias may affect these data. Nonetheless, this dataset is consistent with what is usually found, not only in soils but also in many other ecosystems where kingdom *bacteria* includes more lineages than others. *Archaea* is commonly the least prominent and the least diverse. The lower diversity of *fungi* is fully expected due to their comparatively recent evolution (Hug et al., 2016). In soil ecosystems *bacteria* richness is expected to be 2-3 times larger than *fungi* species (Peay et al., 2016).

In our dataset, we found more samples hosting *fungi* (247 of 294) than *bacteria* (228 of 294) and/or *archaea* (77 of 294). In general, *fungi* and *bacteria* were traced almost all over the sampling sites along a $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$ - excepting sampling site #1, which was absent of *bacteria* observations most likely, due to a methodological error -. Instead, *archaea* were found only within 33 of the 49 sampling sites along both transects. Both *bacteria* and *fungi* can be analysed at the same sampling effort (10,000 sequences) given the robustness in comparative evaluations between these two groups of microbes. In such case, Peay et al., (2016) pointed out to keep in consideration that *bacteria* (based on 16rSSU marker) tends to group species together and thus, in contrast with ITS marker, emphasizes higher-level taxa (genera or families) with broader geographical distributions, possibly but not necessarily, reflecting conserved physiological features.

Microbial diversity metric calculation and taxonomic assignment

On the basis of the DNA sequencing dataset described above, we estimated richness, evenness and taxonomic composition. Soil microbial *α-diversity* and *β-diversity* and taxonomic analyses were made using QIIME (Caporaso et al., 2010b) and the R package "Phyloseq" (McMurdie and Holmes, 2013)– described in Chapter 3. Below we detail the workflow and specifications applied during these analyses. Some of the settings and other additional features were determined by analysing the descriptive statistics of the microbial DNA sequencing datasets described above and also shown in Chapter 3 (*e.g.* number of sequences per OTU/sample).

In QIIME, diversity analyses were run on *bacteria*, *archaea*, *unassigned prokaryotes* and *fungi* OTUs datasets (*i.e.* .biom tables), separately. Each of these microbial data were called through a core set of diversity scripts involving *α-diversity*, *β-diversity* and taxonomic analyses altogether (*i.e.* core_diversity_analyses.py script). For all these taxa, *α-diversity* in terms of richness was measured using Chao1, observed OTU's and Phylogenetic Diversity (PD) indices; whereas, *α-diversity* evenness was quantified by Shannon index (H') and Simpson's index (D) indices. On its own, *β-diversity* richness was measured using Jaccard[xxxi], Sørensen[xxxii], and Unifrac indices; meanwhile, *β-diversity* evenness was quantified using Bray-Curtis[xxxiii], Morisita-Horn[xxxiv] and

[xxxi, ii] *Measure of communities similarity.*

[xxxiii,v] *Measure of communities dissimilarity.*

Weighted Unifrac indices.

The sampling efforts for these diversity calculations were set up differently for each taxa group at: 80 (*archaea*); 1,000 (*unassigned prokaryotes*) and 10,000 (*bacteria* and *fungi*) sequences. Such values were determined based on: (i) number of sequences/sample; (ii) number sequences/ecosystems for maintaining all the ecosystems (cropping /natural) and sampling sites included; and (iii) rarefaction curves plateau. The rarefaction was incorporated by QIIME diversity pipeline in both a single and multiple rarefaction processing, by which, all the soil samples were equalled to the same sequences sampling effort.

Diversity results for each metric for the three kingdoms are shown in Appendix 2 organized by transects and ecosystems (Table 5-1 to Table 5-4). Richness diversity results for NS-$_{transect}$ and WE-$_{transect}$ are shown in Table 5-1 and Table 5-2, respectively. Evenness diversity results for NS-$_{transect}$ and WE-$_{transect}$ are shown in Table 5-3 and Table 5-4, respectively.

MICROBIAL TAXONOMIC COMPOSITION

The taxonomic distribution across the sampling sites and both type of land uses is described below for each microbial kingdom. At phylum level, we found 48 types of *bacteria*, 3 types of *archaea,* and 7 types of *fungi*. On a basis of the relative abundances, 87.7% of *bacteria*, 97.1% of *archaea* 94.1% of *fungi*, are reserved by the first dominant phyla (>5%) (Figure 4-1). The relative abundance of those most abundant phyla within each kingdom were not significantly different in both types of land use. Therefore, the taxonomic composition of the most abundant phyla were almost the same across all habitats but slightly variable in term of their abundances - mainly in *fungi* communities (Figure 4-2). For example, the most abundant *bacteria* and *archaea* (>5%) concentrated 88% *vs* 87% and 97% *vs* 98% of the individuals in natural vs cropping ecosystems, respectively. By itself, the three most abundant *fungi* phyla concentrated 96% in natural ecosystems and 92% in cropping ecosystems. In all kingdoms, the foremost differentiations related to the taxonomic composition by land use were attributed to rare communities, either because of the incidence of certain rare types (unshared taxa), *e.g*. *Fusobacteria* only found in cropping ecosystems; or because of differences in the relative abundance of certain small groups (Figure 4-3) – on the basis of our literature review in Chapter 1 about the advantageous of amplicon approach to identify rare communities is that we considered them important in this analysis. For example, although natural ecosystems concentrated a higher abundance of rare individuals than cropping ecosystems (see Figure 4-11), we found slightly larger proportion of rare taxa in cropping ecosystems - dissimilarity that was also more evident in *fungi* (9% vs 4%)-. In addition, another particularity of *fungi* was that all the unshared taxa belonged to the most

abundant phylum *Ascomycota* (*e.g. Dothideomycetes*, *Eurotiomycetes* classes). In general at the genus level, both natural and cropping ecosystems shared 86% of *bacteria*, 75.6% of *fungi* and 61.5% of *archaea*. More details about taxonomic differences between ecosystems are discussed below.

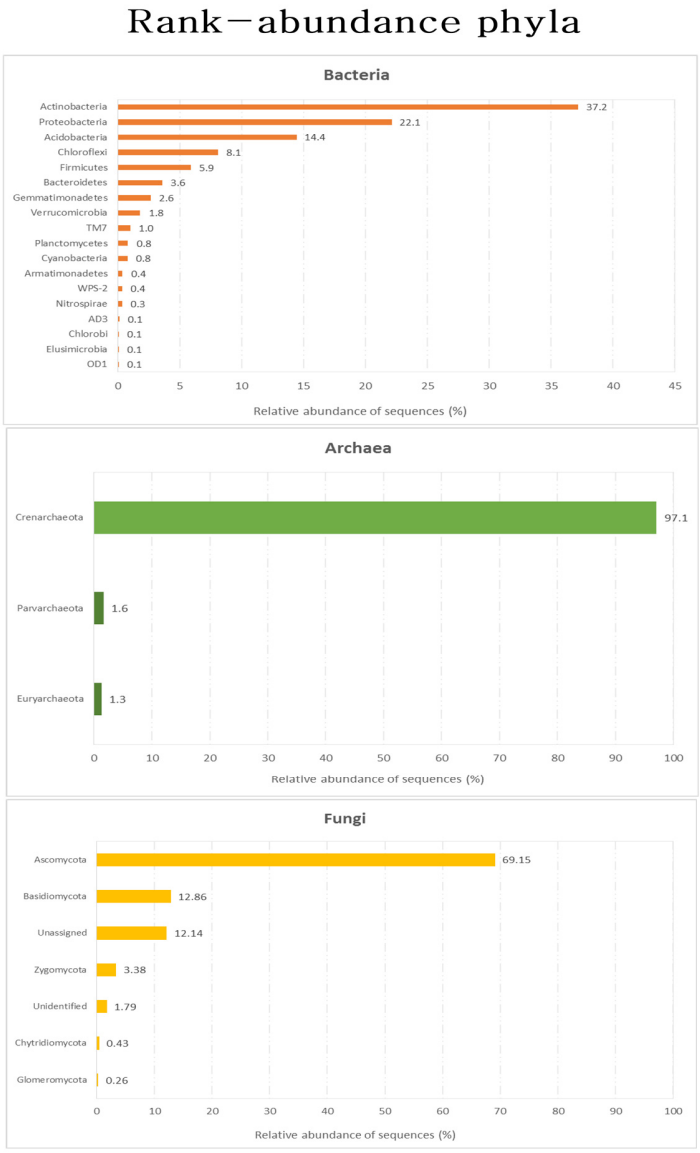


Figure 4−1. Relative abundance of phyla found in the entire study area. In *bacteria*, there were found 48 phyla (L2) and five of them represented the most abundant (>5%) by concentrating 87.7%. In *archaea*, there were observed 3 major phyla (L2) in which Crenarchaeota phylum concentrated 97.1%. *Fungi* were 7 major phyla (L2) and 2 of them remained undefined. From these 7 phyla, 3 most abundant (>5%) concentrated 94.1%.

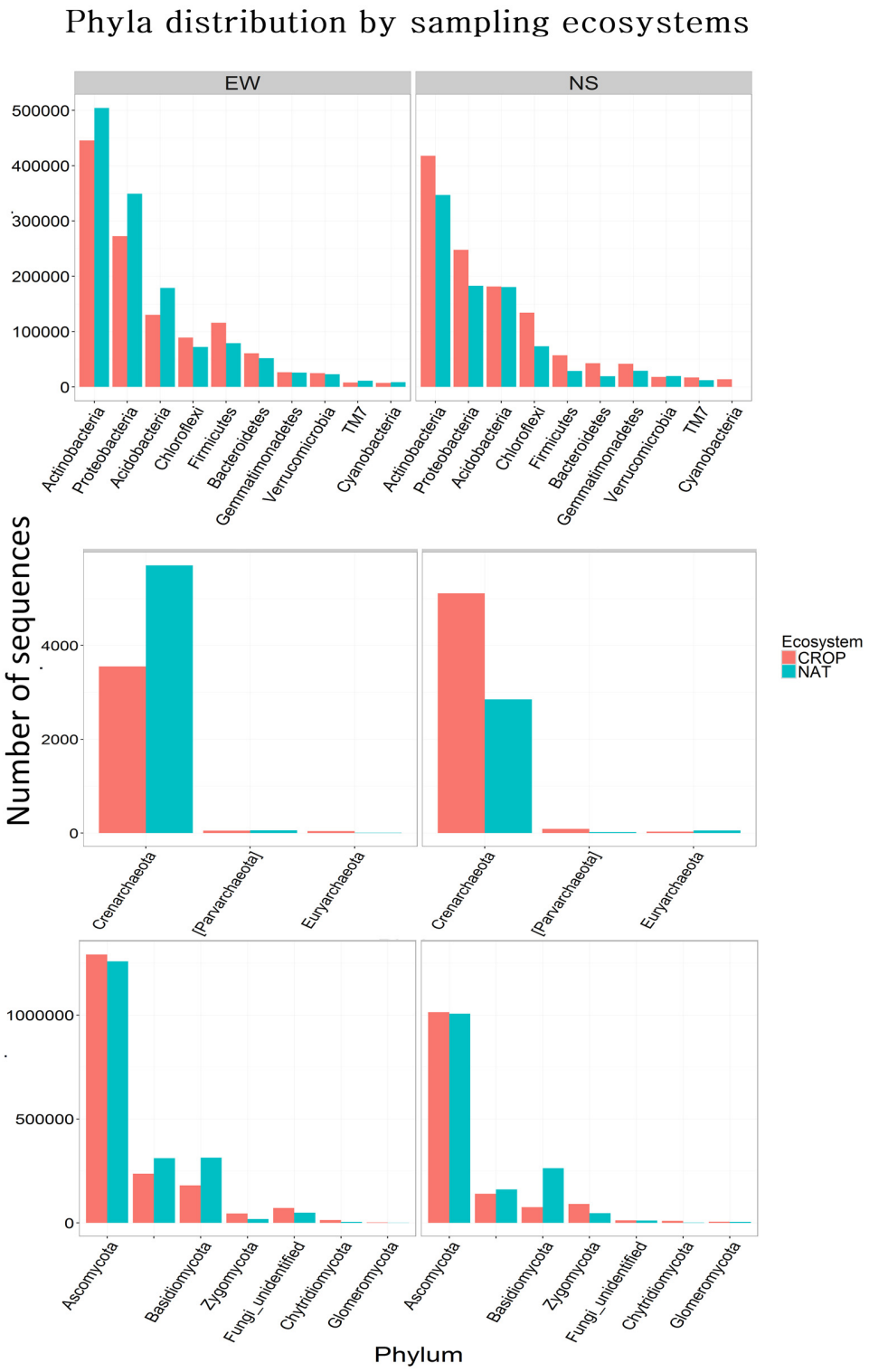


Figure 4-2. Relative abundances of the most abundant phyla within the different agroecological environments sampled across NSW.

Bacteria

The most abundant *bacteria* we found were *Actinobacteria* (37.2%), *Proteobacteria* (22.1%), *Acidobacteria* (14.4%), *Chloroflexi* (8.1%) and *Firmicutes* (5.9%). On the other hand, the lower abundant phyla (<5%) were *Bacteroidetes* (3.5%), *Gemmatimonadetes* (2.7%), *Verrucomicrobia* (1.8%); and the most rare phyla (<1%) were *TM7* (1.0%), *Planctomycetes* (0.8%), *Cyanobacterias* (0.8%), *Armatimonadetes* (0.4%), *WPS-2* (0.4%), *Nitrospirae* (0.3%), *AD3* (0.1%), *Chlorobi* (0.1%), *Elusimicrobia* (0.1%), *OD1* (0.1%), among 30 others (Figure 4-3).

The phylum of *Actinobacteria* are recognized for being widely distributed in both terrestrial and aquatic environments, yet are especially abundant in soil habitats where they play a crucial role in the decomposition of biomaterials and in the humus formation (Valverde et al., 2012; Ventura et al., 2007). At the genus level, *Rubrobacter* (8%) from *Actinobacteria* phylum, and *Bacillus* (3.6%) from *Firmicutes* phylum, were the most abundant *taxa* we found. The other abundant group of *Proteobacteria,* is recognised for being the largest and, phenotypically, most diverse phylum within the *Bacteria* kingdom thus far. The ecological contribution of this group is highly recognized in innumerable biogeochemical processes, which most probably relies on their global distribution and environmental preferences (Spain et al., 2009). *Rhizobia*, one of the most iconic contributor to plant nutrition by fixing nitrogen gas when living in symbiosis with leguminous plants belongs to this group of *bacteria* (*Alphaproteobacteria*) (Gu and Mitchell, 2006).

Regarding the effect of land use within the most abundant *bacteria* phyla, *Chloroflexi* marked the most significant difference between the *bacteria*l composition of cropping and natural land uses

(9.3% vs 6.5%, respectively). Marginally, *Gouta4* and *Synergistetes* phyla were exclusively observed in one or two natural sites. Conversely, *Fusobacteria*, *GNO4*, *NC10* and *SR1* phyla were only found in few cropping/disturbed ecosystems (Figure 4-3). The taxonomic bacterial distribution per sampling site along the $NS_{-transect}$ and $WE_{-transect}$ is shown in Figure 4-4.

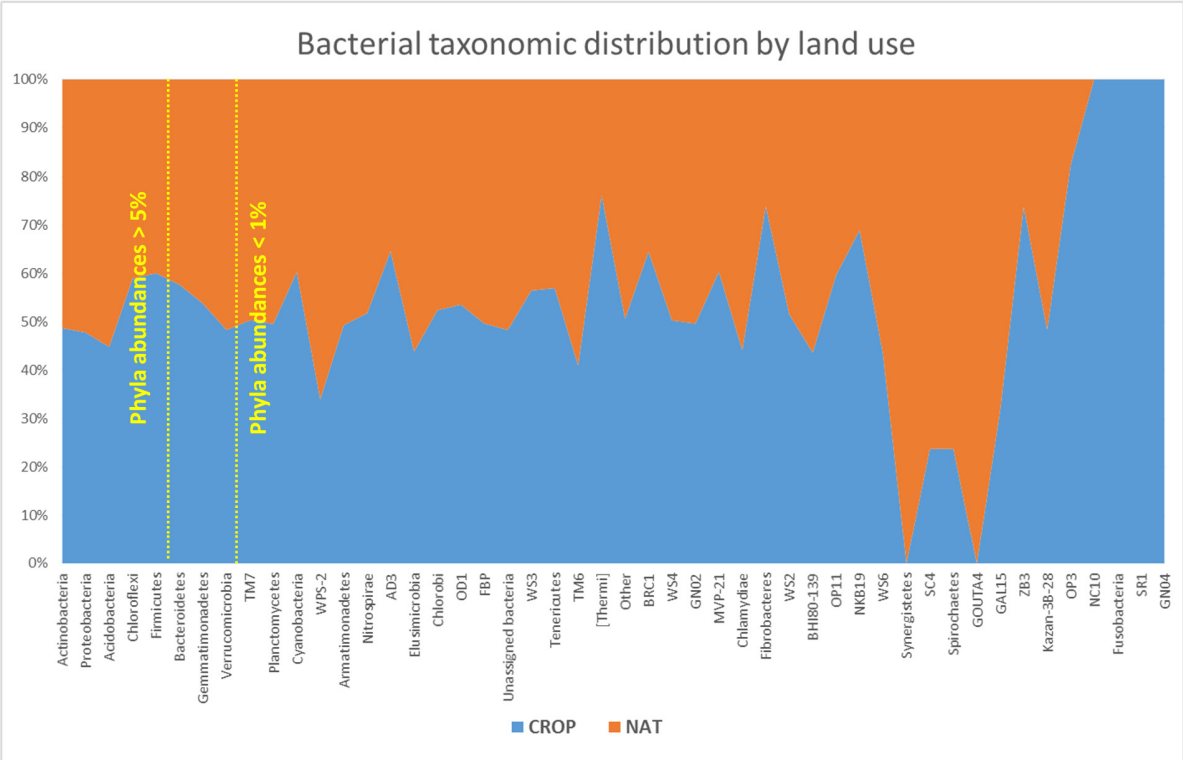


Figure 4–3. Bacterial taxonomic distribution at phylum level in both natural and managed ecosystems.

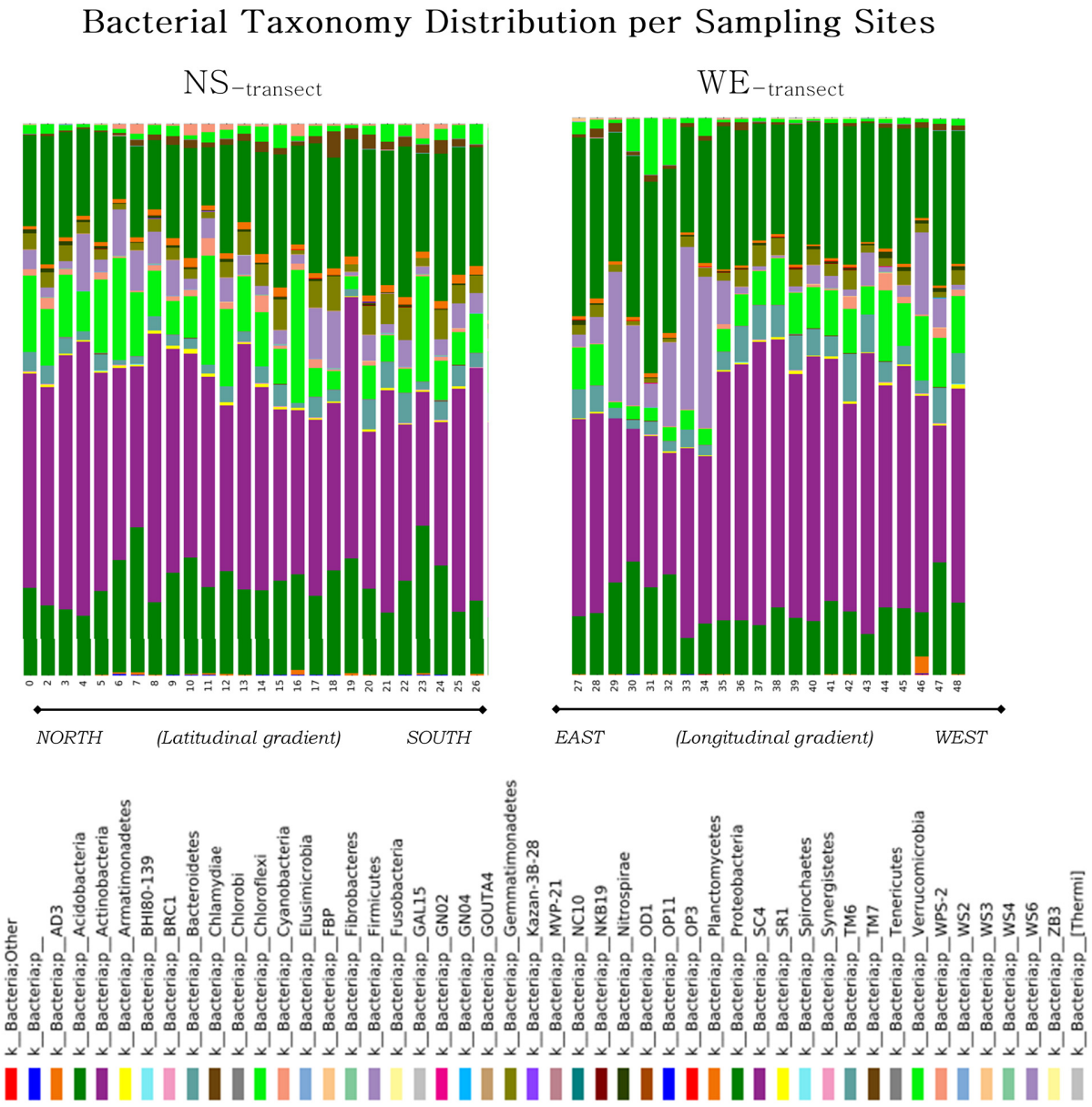


Figure 4−4. *Bacteria* phyla distribution across the sampling sites.

Archaea

In *archaea*, *Crenarchaeota* (97.1%) was the most dominant phylum. Within this phylum, *Nitrososphaera genus* from *Thaumarchaeota* class occupied *97.8%* of total *archaea* datasets. This group has been widely documented as the most dominant *archaea* (Schleper and Nicol, 2010). *Nitrososphaera* is an ammonia oxidizing microorganism involved in the transformation of ammonia to nitrite in aerobic conditions. In contrast to other anaerobic *archaea* closely related to carbon cycle (*e.g.* methanogenesis and methane oxidation), a preference for aerobic conditions allow this group of *archaea* to be abundant and distributed across different ecosystems, including extreme environments such as hot springs (Offre et al., 2013). The least abundant phyla of *archaea* were *Parvarchaeota* (1.6%) and *Euryarchaeota* (1.3%).

As mentioned above, the differences between natural and managed ecosystems in the taxonomic composition of *archaea* were associated to the rare community. For example, from *Euryarchaeota phyla* an unassigned *genus* of the class *Thermoplasmata* (acidophiles which prosper in high-temperatures environments) was recognized in natural/undisturbed environments and very few times in cropping/disturbed or uncovered areas. Conversely, certain *Methanobacteria* (methane producers in anoxic conditions) appeared only in disturbed environments but none of them were observed in natural sites. The taxonomic distribution of *archaea* phyla per each sampling site along both transects is illustrated in Figure 4-5.

The low diversity and abundance we found in *archaea* in this survey (at 0-5cm depth) may be consistent with the documented fact that these microorganisms are normally more diverse at

depth than in the upper soil surface (Cao et al., 2012; Hartmann et al., 2009; Uroz et al., 2013).

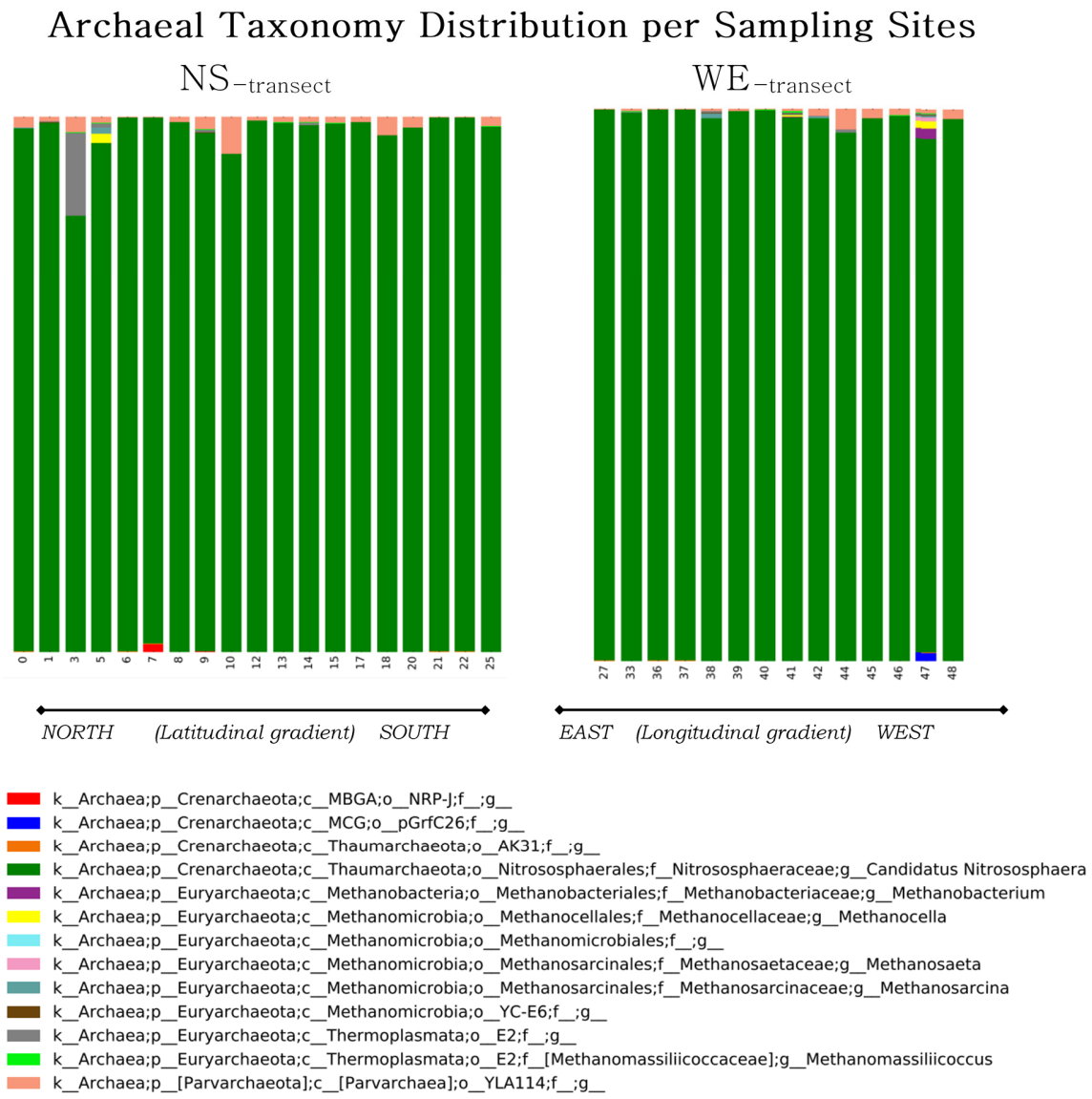


Figure 4-5. Archaeal phyla distribution across the sampling sites.

Fungi

In *fungi*, the most abundant phyla were *Ascomycota* (69.1%), *Basidiomycota* (12.9%) and other undefined (12.1%) phylum. Conversely, the less abundant phyla (<5%) were *Zygomycota* (3.4%), another *unidentified* group (1.8%) and *Glomeromycota* (0.3%). The most abundant *Basidiomycota* and *Ascomycota* are groups largely recognized for being vastly abundant in soils, in fact, more than in any other ecosystem (Peay et al., 2016). Both phyla are well- known as the primary litter-decomposers organisms in the upper layers of soils. *Basidiomycota* are among the few organism able to degrade residual plant biopolymers such as lignin, by which, they play an important role in the decomposition of organic matter, mainly in forest soils (Kuramae et al., 2013; Ma et al., 2013). On the other hand, *Ascomycota* seem to be primary decomposers in agricultural soils whose crop residues have lower content of lignin (Kuramae et al., 2013). In any case, there is also confirmed an important relationship between both groups during the organic materials decomposition on which *Basidiomycota* increase over time whilst *Ascomycota* decrease (Kuramae et al., 2013). From the most abundant phylum *Ascomycota*, *Fusarium* (3.1%) and *Penicillium* (2.5%) were the most abundant *genera* in our study area.

The most significant *fungi* variation between land uses was given by phylum *Basidiomycota* whose population in natural ecosystem is more than double the amount in managed sites (8% *vs* 17%) (Figure 4-9, Appendix 1). This difference was mostly attributed to *Agaromycete* class (15% vs 5.3%), which includes important groups of root symbionts associated to support the growth of exotic Australian acacias in forested areas (*e.g. Thelephorales* and *Glomerales* families) (Bâ et al., 2010). Similarly, Lauber et al., (2008) found greater abundances of *Agaricales* when comparing

forest (hardwood and pines) *versus* cultivated/livestock pasture lands of north-western South Carolina, USA. In this case, the author attributed this result to the differences in the soil nutrient status (*e.g.* soil C: N ratio and extractable P).

C. MICROBIAL DIVERSITY MEASUREMENTS

Richness and evenness

a. Completeness of sampling effort

The first insight in α-diversity at each location can be seen through the rarefaction curves obtained for each diversity metric performed by QIIME (Caporaso et al., 2010b). These accumulation curves provided an overview to the richness and evenness of the microbial structure at each of the sampling sites in our study area. Additionally, these plots provided an extent on how this diversity pattern increases with sampling size and sample completeness. In this regard, all of the rarefaction curves vary according with the diversity metrics calculated and there is a disparity when scoping the total richness represented at the different level of sampling. However, there is not significant variations across the curves when analysing the diversity pattern per sampling site by the different metrics and a close and similar trend is maintained.

For instance, the rarefaction curves based on *observed OTUs* per sampling sites are shown for each taxon in Figure 4-6. In these plots, the completeness of sampling effort at 10,000 seqs for *fungi* community seems higher than for bacterial community whose rarefaction curves show to be more distant from reaching a plateau. The small group of *archaea* seems to be quite well

sampled at 80 seqs but *unassigned prokaryotes* is even the least asymptotic sampled at 1,000 seqs.

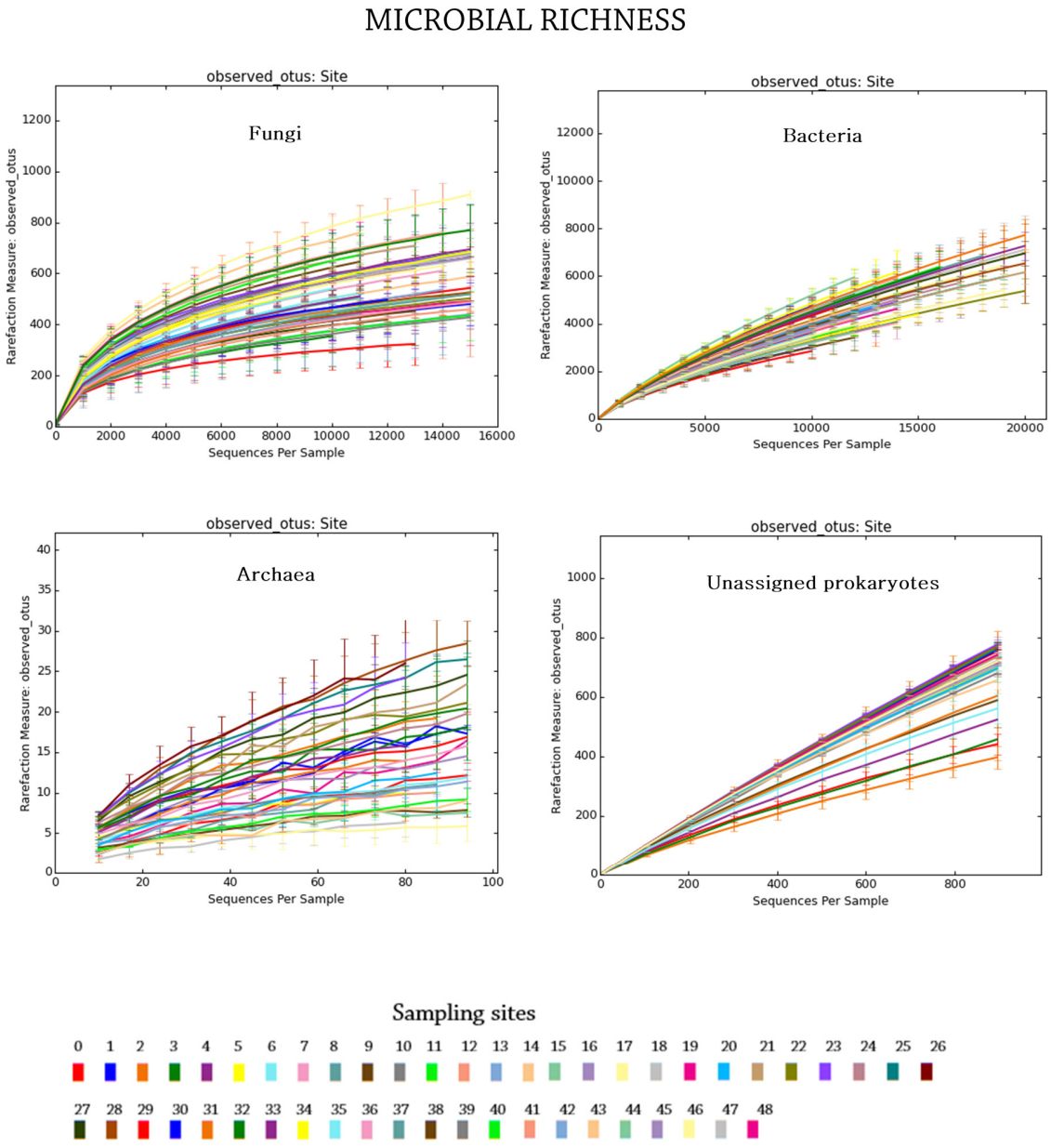


Figure 4–6. Rarefaction curves showing richness accumulated in terms of the observed OTUs per sampling sites in the entire study area. Each curve displays the result for the mean from 10 iterations of the data.

In spite of these results, we have decided to keep this level of sequencing effort for downstream analysis instead of increasing the cut-off and hence, sacrifice the number of samples and even some of the sampling sites. Indeed, there were found significantly fewer sampling sites curves represented at the level of 15,000 and 20,000 seqs for *fungi* and *bacteria*, respectively. What is more, these results were slightly different when analysing the rarefaction curves on the basis of other richness and evenness estimators such as *Chao1*, *Phylogenetic Diversity* (PD) or *Shannon index* (H'). For instance, the rarefaction curves shows for *Chao1* and *PD* metrics (Figure 4-7) reveals to be more completely sampled when characterizing *bacteria*'s richness at 10,000 seqs than the observed OTUs shown above. The sampling effort has been significantly influenced by abundance information in all of the microbial data. For example, bacterial plots are undoubtedly asymptotic curves for *Simpson* and *Shannon* measurements (Figure 4-7). The resulting differences along diversity metrics supported the essential of testing biodiversity using more than one single diversity metric because their vulnerability to sampling size is variable (Morris et al., 2014).

b. Richness and evenness discrepancies among microbial kingdoms

Richness discrepancies among microbial kingdoms among sampling sites are evident in these rarefaction plots (Figure 4-6). We observe that both richness levels and the degree of differences among the 49 sampling sites are significantly different. This became clearer as sampling efforts increase in all the cases. In this sense, the average from 10 iterations at a depth of 10,000 sequences has ranked the richest sampling site up to ~ 5,500 different types of *bacteria,* whereas the richest site for *fungi* contained no more than ~ 800 different types. The small group of *archaea* comprised no more than ~ 30 different OTUs in the *archaea's* richest sampling locations.

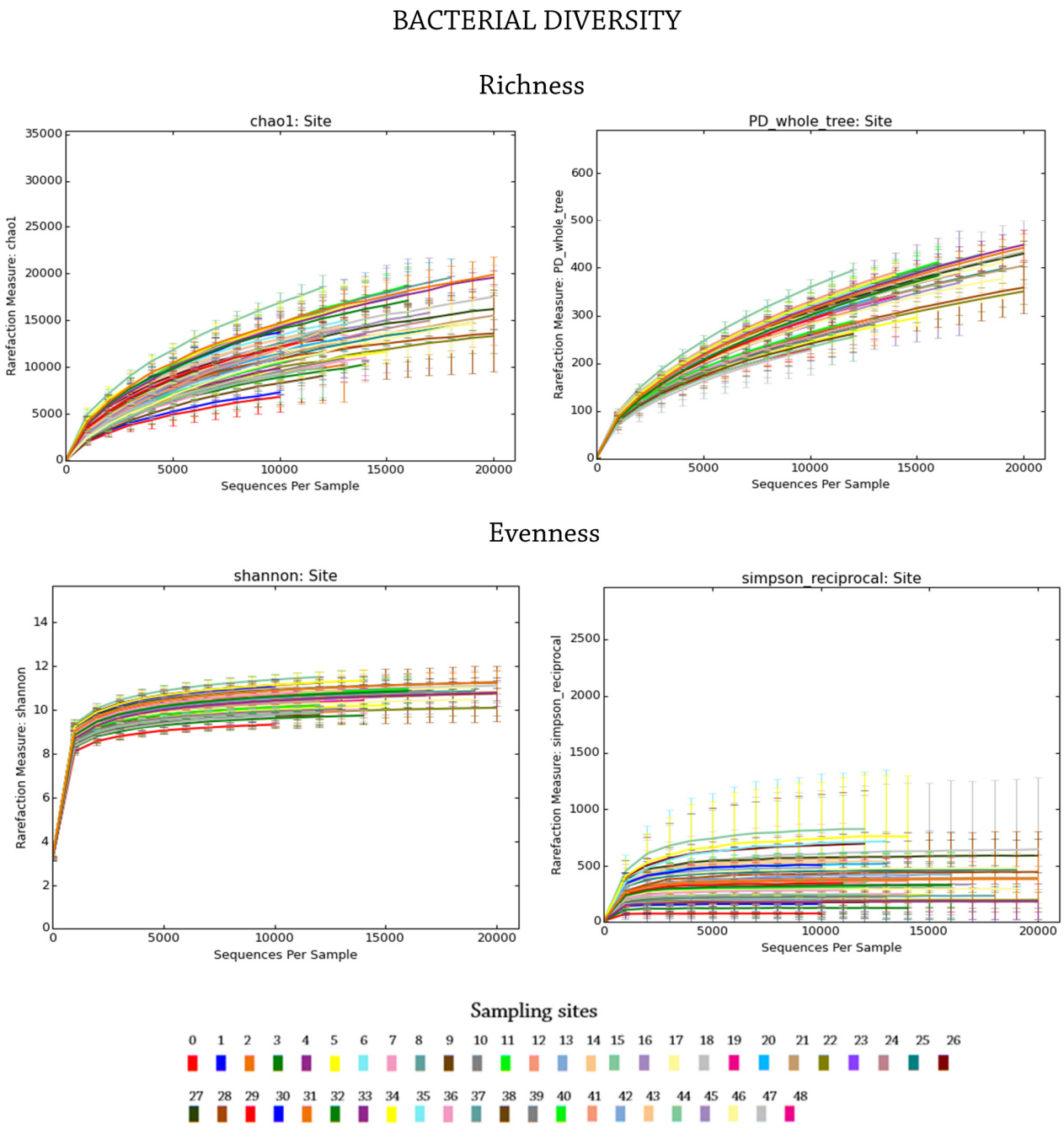


Figure 4−7. Bacterial diversity characterised using both richness and evenness estimators.

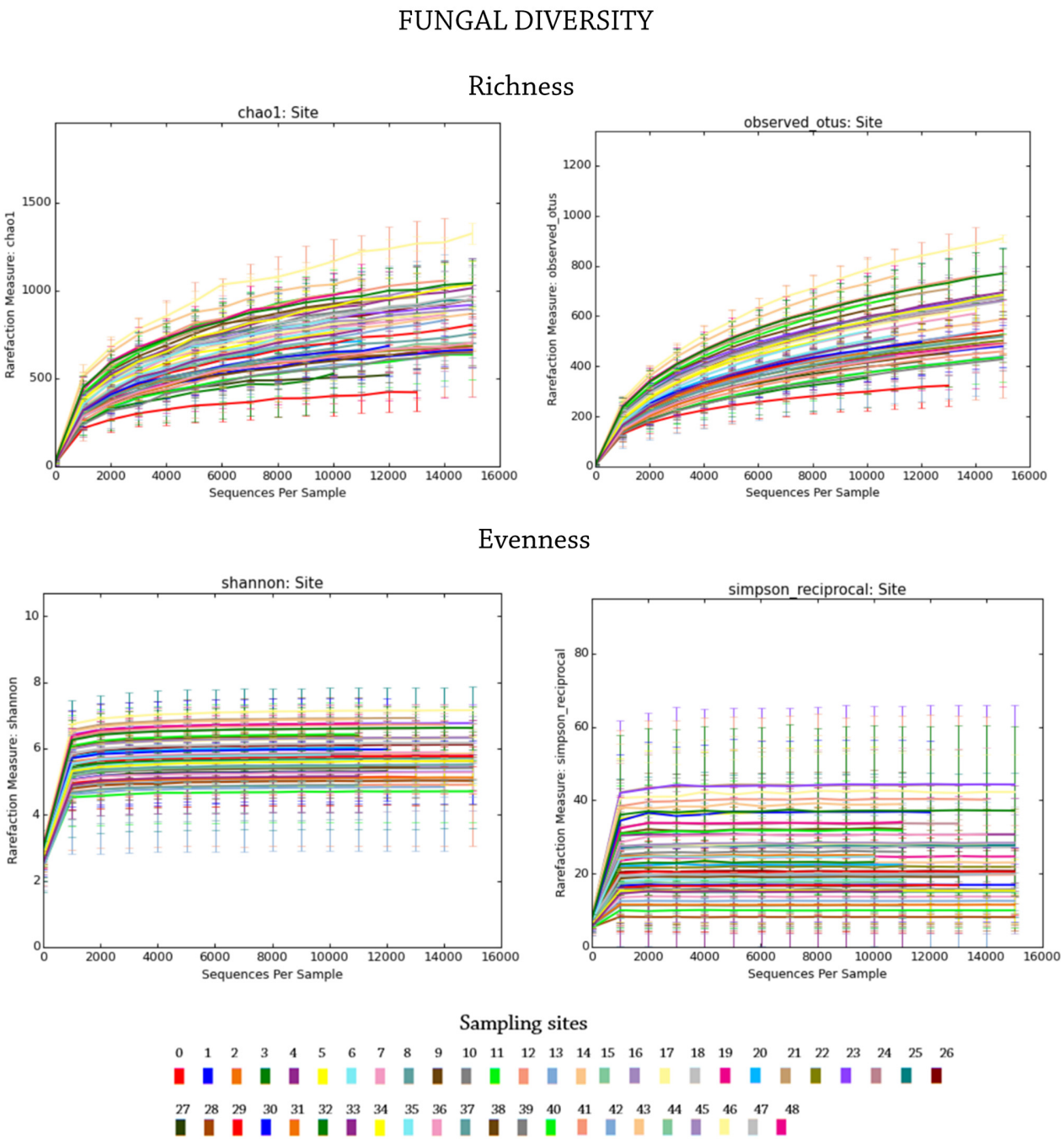


Figure 4−8. Fungal diversity characterised using both richness and evenness estimators.

Similarly, bacterial richness measured by *Chao1* ranged nearly 4,000 to 17,500 different species, whereas, *fungi* richness was approximately between 270-1200 species. On its own, *archaea* presented *Chao1* values oscillating between 6-55 species. The richness analysis of prokaryotes by *Phylogenetic Diversity* (*PD*) metric has shown less dispersion than observed OTUs among the sampling sites curves (Figure 4-7). This result might indicate that prokaryotic communities exhibit more diversity variations from an evolutionary perspective than from a taxonomic point of view. This indicates that a variation at the phylogenetic lineage was not visible at the taxonomic categories (e.g. phyla or genera). In other words, the phylogenetic relatedness of taxa is more variable across prokaryotic communities but taxonomic binning methods may fail to detect this variation or were sensitive to the choice of threshold for identifying distinct taxa. The specific diversity measurements for each sampling sites obtained from the different diversity metrics are specified in the Table 5-2 and Table 5-1 (Appendix 1) for $NS_{\text{-transect}}$ and $WE_{\text{-transect}}$, respectively.

A first insight suggested a more even distribution of *fungi* than *bacteria* communities as the fungal accumulation curve seems steeper than *bacteria*'s curve in term of the amount of observed OTUs (Figure 4-6). On the other hand, when estimating diversity based on evenness estimators such as *Shannon (H')* and *Simpson (1-D)* indices, the group of *bacteria* were more evenly distributed in most of the sampling sites (H'= 9.32-11.61 ; 1-D > 0.98) than *fungi* group (H'= 2.96-7.23 ; 1-D > 0.67). This evidence suggested that by reducing the sensitiveness by richness in diversity calculations, the bacterial community turned to be slightly more equally distributed than *fungi community*. Likely, this indicates a higher influence of the rare communities in the

prokaryotic group principal contributors to increment the total richness. For example, the rare communities affecting an evenly distributed pattern can be seen in the increase of the steepness when changing from richness to evenness plots, which has been observed for all the taxa (Figure 4-7 and Figure 4-8).

These results were also corroborated by estimating specific evenness measurements using both *Simpson* and *Shannon* metrics. Both evenness metrics are constrained between 0 and 1. *Shannon evenness (i.e. J'=H'/H_{max};* Pielou, 1975) has observed bacterial (0.82≤J'≤0.9) community more evenly distributed than *fungi* (0.41≤J'≤1.0) and archaeal (0.1≤J'≤0.65) communities. Additionally, this *bacteria* evenness level was less variable among the sampling sites. However, this metric is more sensitive to richness, which is significantly larger in *bacteria* (Magurran, 2004) – by itself this index is mathematically correlated to H' (Morris et al., 2014). Then, in a way to reduce the richness component for comparison, we also measured Simpson evenness index (*i.e.* E1/D =(1/D)/Obs_OTUs; Smith et al., 1996) – which advantage relies on its total independence from *Simpson diversity* (1-D) (Morris et al., 2014). In this case, the resulted evenness values fluctuated ~0.02 ≤ $E_{1/D}$ ≤ 0.24 in *bacteria*; 0.01 ≤ $E_{1/D}$ ≤ 0.14 in *fungi* and ~ 0.13 ≤ $E_{1/D}$ ≤0.49 in *archaea*. Accordingly using this metric, *archaea* and *bacteria* were more evenly distributed but their values were also more disperse among the sampling sites. By contrast, *fungi* were more uneven but this pattern is more constant among sampling sites. These patterns can be easily influenced by the high abundances of *fungi* communities, which was significantly higher in term of sequences.

Indeed, the kingdom *fungi* presented more dominance/commonness than *bacteria* and *archaea* and, this dominant pattern was observed in most of the sampling sites. Even though *archaea* and *bacteria* are more uniformly distributed, they displayed more rareness levels than *fungi*. This even distribution pattern, however, was more variable among the sampling sites. Furthermore, $E_{1/D}$ ranges pointed out that by reducing the richness component, and although *bacteria* were shown as more uniform compared with *fungi*, all of these microbial kingdoms can be considered away from being evenly distributed (values < 0.5).

c. Microbial richness and evenness affected by land use

A global view of the study area revealed an effect of land use in the microbial diversity where natural/undisturbed ecosystems have resulted of higher diversity than managed/cropping ecosystems (Figure 4-9). This pattern was followed by every taxa for almost all the diversity metrics estimated. Only *Simpson (1-D)* diversity showed to almost equal both ecosystems (Table 5-3 and Table 5-4). Reasonably, this result might be related to the influence of dominant microbes that are significantly abundant in cropping ecosystems which is particularly emphasised by this index (Figure 4-11).

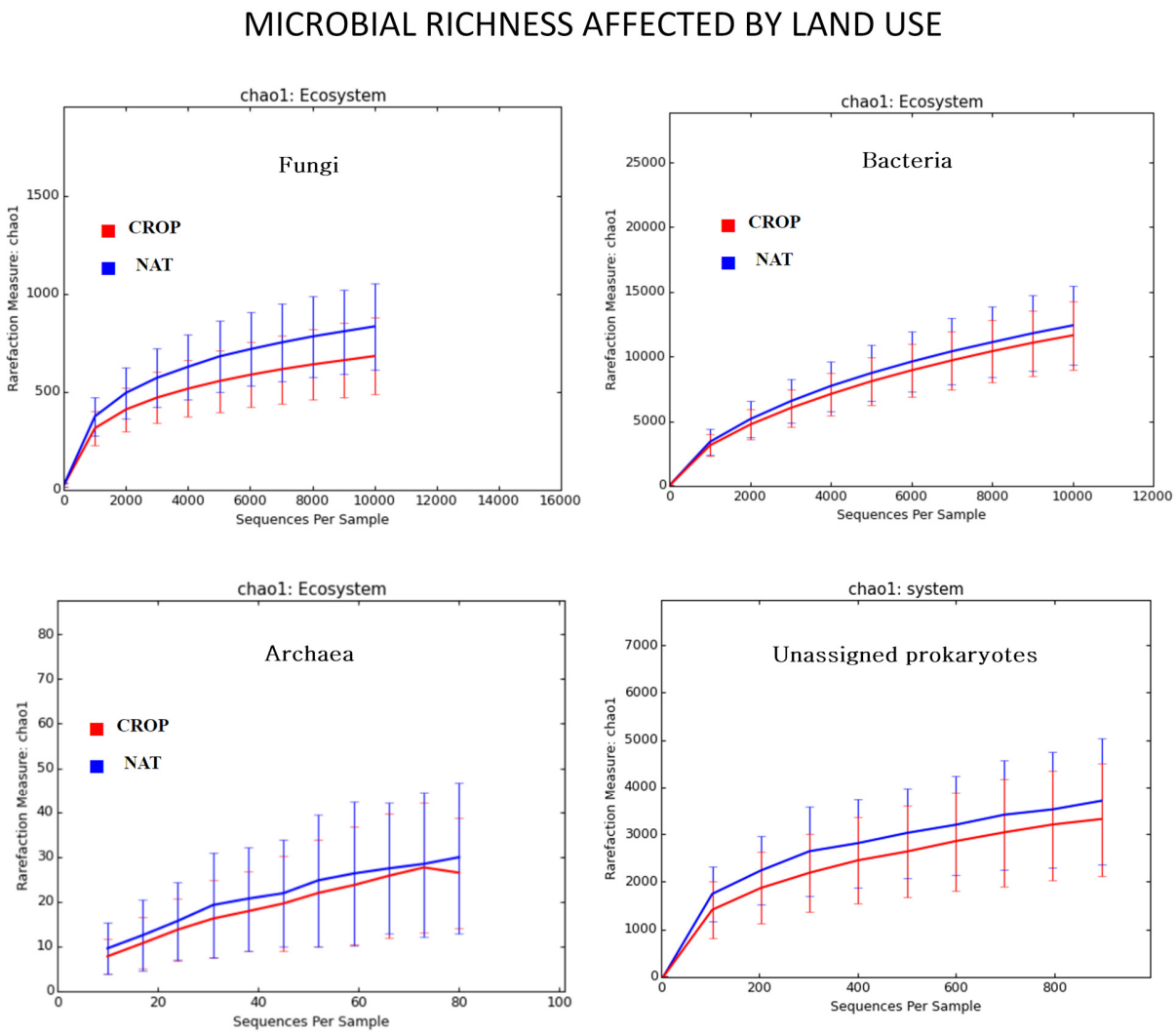


Figure 4−9. Microbial diversity affected by land use. It is shown the richness ranges reached at the two ecosystems under study: natural or undisturbed vs cropping ecosystems. This same tendency was seen all across the other diversity richness and evenness estimators.

Microbial abundance: commonness and rarity

d. Commonness and rarity discrepancies among microbial kingdoms

Complementing diversity analyses, we estimate other microbial diversity patterns on the basis of graphic analyses using 'OTU-accumulation curves' and 'OTU rank-abundance curves'. In these cases, 'OTUS-accumulation curves' are represented by the rarefaction curves plotted by QIIME as part of the multiple rarefaction procedure. Then, 'OTU rank-abundance curves' were plotted using observed OTU's data in R "ggplot2" package (Wickham, 2009). Together accumulation/rank-abundance plots and diversity indices were used to analyse *α-diversity* and *β-diversity* along the environmental latitudinal/longitudinal gradients in NSW. Particular emphasis on these analyses was to elucidate soil microbial distribution and soil physicochemical properties relatedness.

The rank-abundance of each kingdom is shown graphed in Figure 4-10 to Figure 4-12. Even though the most dominant OTUs represented no more than 1.3%, 4.2% and 36.5%, respectively, of *bacteria*, *fungi* and *archaea* dataset, the order of magnitude of the ranks-abundance plots evidenced the great differences among these three kingdoms. This result highlight the great magnitude of soil microbial diversity, especially, soil *bacteria*. This richness is largely weighted by the rare community represented by the long-tails of these graphs (which include rare OTUs of 3 seqs). *Fungi* drawn more dominant OTUs than *bacteria*. This pattern is demonstrated in the steepness of these plots, in which, *archaea* presented the major degree of dominance followed by *fungi* and then *bacteria* that reached earlier the 0.1% threshold of the rare community.

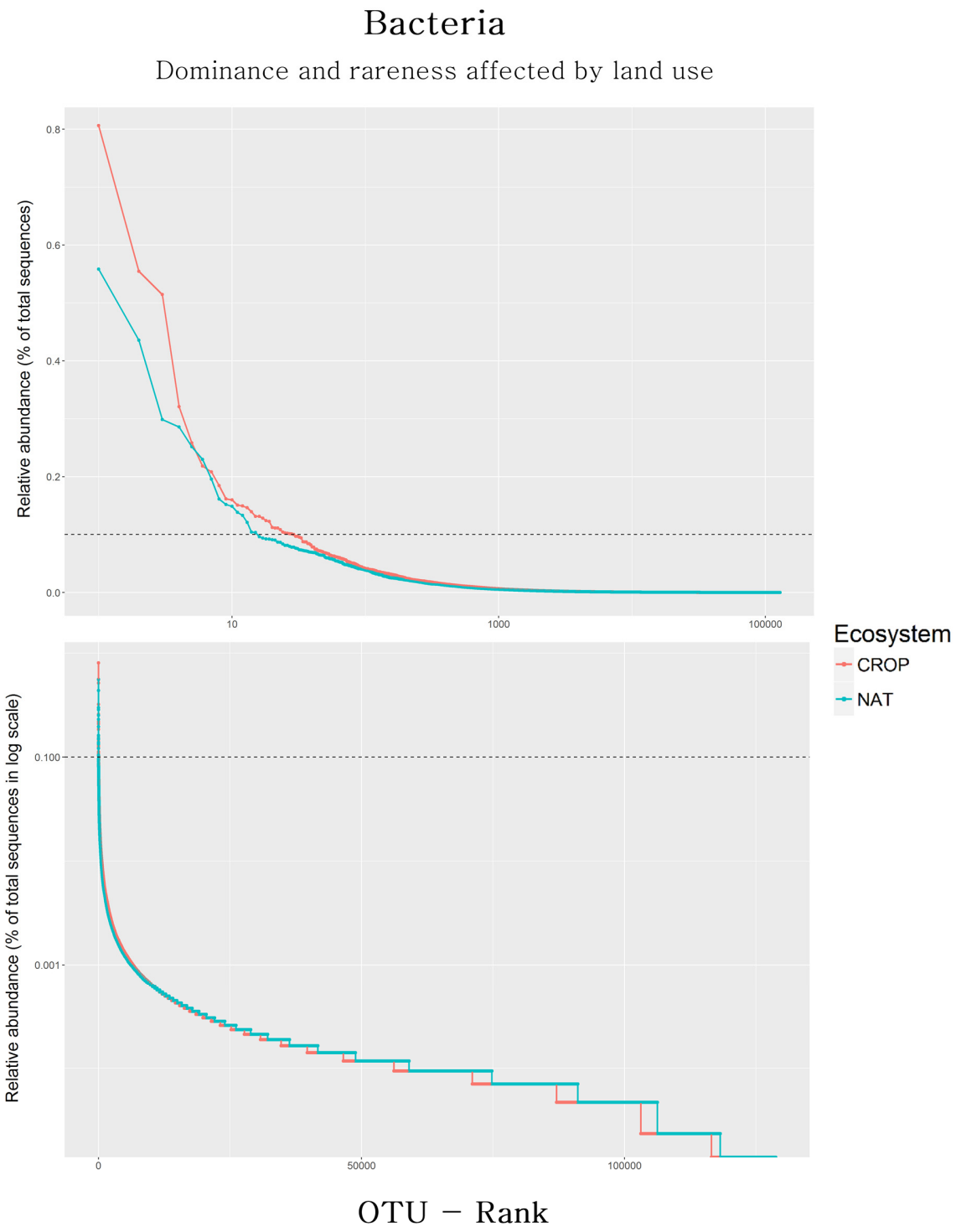


Figure 4−10. Rank−abundance *bacteria*. Relative abundance is shown plotted by ecosystem. The rare community is indicated < 0.1% (dotted line). The top graph plotted OTUs rank in log scale to zoom into the most common community. The graph at the bottom plotted the relative abundance in log scale to zoom into the rare community.

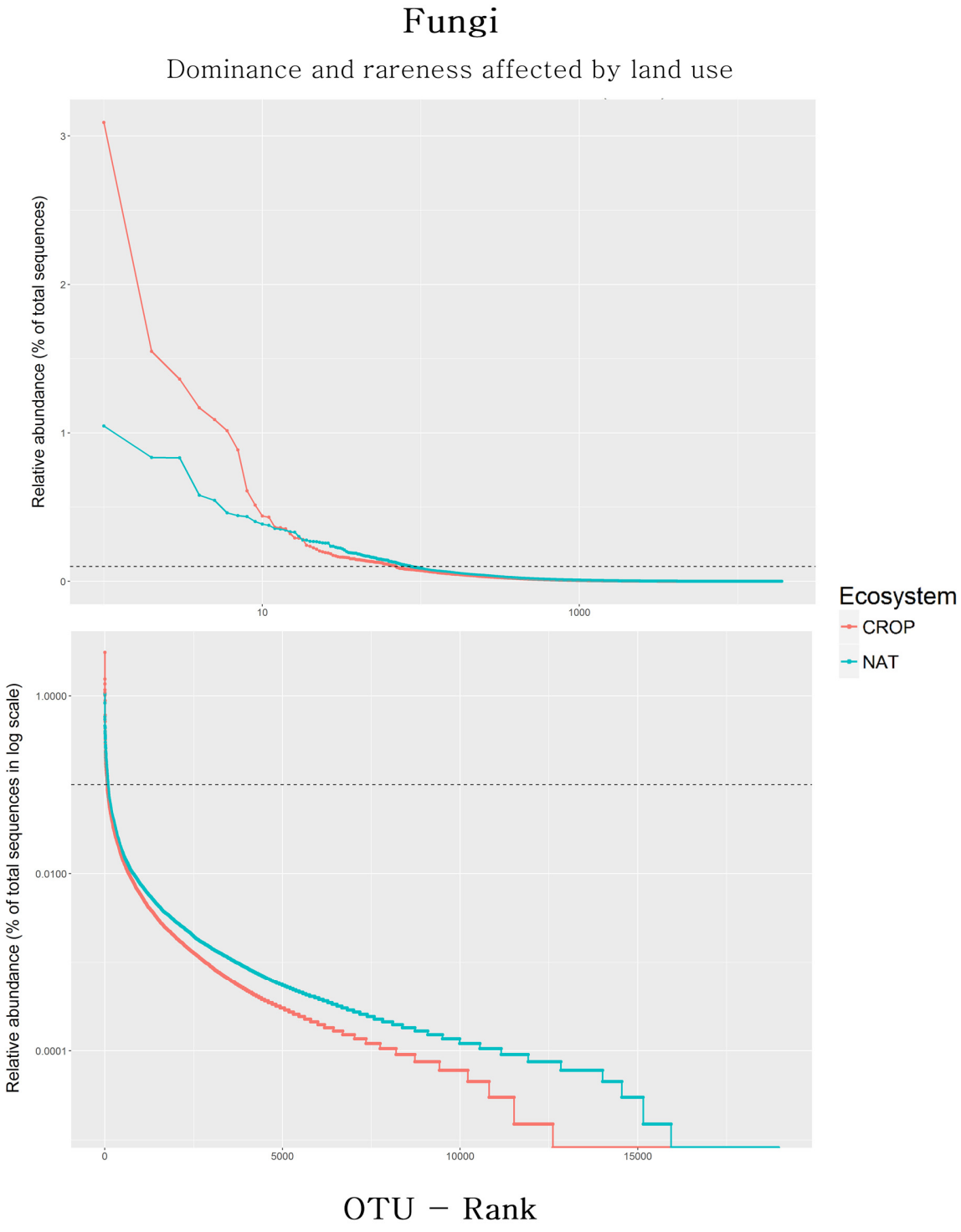


Figure 4−11. Rank−abundance *fungi*. Relative abundance is shown plotted by ecosystem. The rare community is indicated < 0.1% (dotted line). The top graph plotted OTUs rank in log scale to zoom into the most common community. The graph at the bottom plotted the relative abundance in log scale to zoom into the rare community.

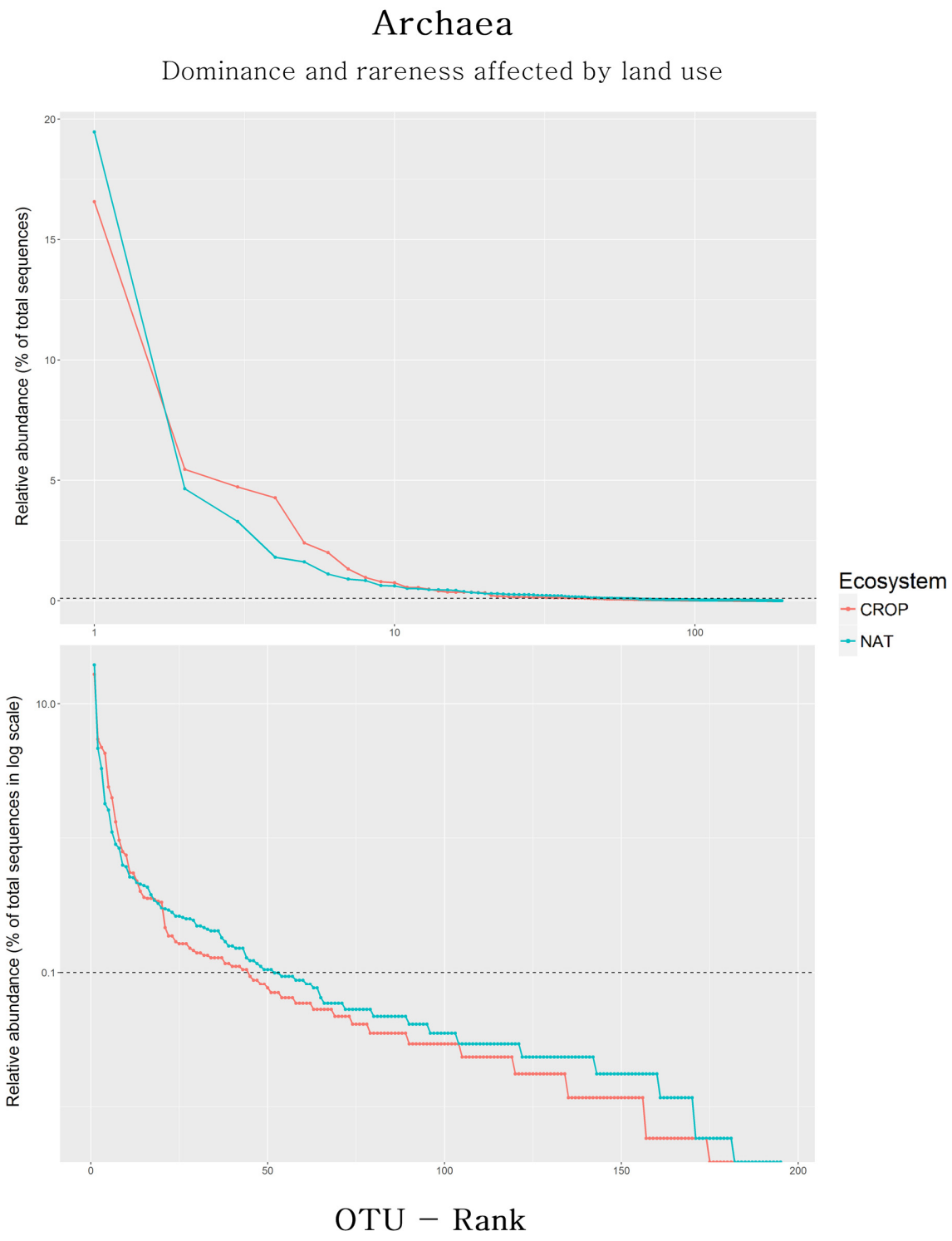


Figure 4−12. Rank−abundance *archaea*. Relative abundance is shown plotted by ecosystem. The rare community is indicated < 0.1% (dotted line). The top graph plotted OTUs rank in log scale to zoom into the most common community. The graph at the bottom plotted the relative abundance in log scale to zoom into the rare community.

Microbial abundance affected by land use

The effect of land use upon soil microbial abundance is also reflected in the rank-abundance plots. There were found more important effect of land uses on dominant OTUs than in the least abundant ones. It is observed how the distance separation between these curves (CROP vs NAT) is reduced toward the rare community. In this regard, the most common *bacteria* and *fungi* species tended to be more abundant in cropping/disturbed ecosystems compared with natural ecosystems. However, in rare communities, this scenario was completely the opposite and the most abundant *archaea* tended to be more abundant and diverse in natural or undisturbed ecosystems. By contrast, the beginning zone of *archaea* curves, *i.e.* the most abundant OTUs, was slightly higher in natural ecosystems compared with cropping ecosystems. Similar results have been found in other studies. For example, bacterial groups were present in higher abundances in a sugarcane field compared to a natural forest in Brazilian Savanna (Cerrado, Central Brazil) (Rampelotto et al., 2013).

The effect of land use were more meaningful in *fungi* community than in *prokaryotic* communities in both common and rare species. Similarly, other studies have also found *fungi* more sensitive than *bacteria* when comparing natural (*e.g.* forest) and cropping ecosystems (Castañeda et al., 2015; Lauber et al., 2008). Moreover, Kasel et al., (2008) findings evidenced strong influences of land use (*e.g.* native eucalypt forest *vs* unimproved pasture) on fungal community in Australian soils in Central Victoria. According to this author argument, "soil *fungi* community is mainly dependent on the availability of suitable habitat because dispersal propagules are readily available for colonisation after land use change".

D. MICROBIAL DIVERSITY ALONG ENVIRONMENTAL GRADIENT

Microbial α-diversity along latitudinal gradient ($NS_{-transect}$)

One common question in microbial ecology is whether there is an increase in species richness or 'biodiversity' from the poles to the tropics (Fuhrman, 2009; Peay et al., 2016). This pattern is often referred to as the latitudinal diversity gradient (LDG) and it is widely recognised in macroorganisms but still under debate for microorganism. In this particular survey, only *archaea* had a tendency to increase northerly, *i.e.* towards the *Tropic of Capricorn (e.g.* H'=2.8 to 0.7 from north to south; Figure 4-15). However, *fungi* and *bacteria* showed an opposite trend increasing their diversity with latitude, *i.e.* from the towns of *Mungindi* to *Howlong* (Site#0 to Site#26; ~900 km distance). Indeed, *bacteria* and *fungi* diversity ranged H'=9.5 to 10.7 and H'=5.5 to 6.3, respectively (Figure 4-13 and Figure 4-14*)*. Despite these increasing diversity toward the poles and tropics, our results may be indicating a more subtle biogeographical pattern as we will see further in this chapter.

Especially for *fungi*, our results agreed with a recent revision of Peay et al., (2016), whose stated that a latitudinal gradient of *fungi* diversity is variable depending on the taxon and functional guild that are investigated. According with the reviewed by these authors, built environments use to show *fungi* increases diversity with latitude. In addition, they stated that is more evident in recent years that most of the fungal species do not have a cosmopolitan distribution pattern. In such case, this explanation is not consistent with Fenchel and Finlay (2004) position about

the ubiquity of small species that tend to have a cosmopolitan distribution. From Fenchel and Finlay (2004) point of view, the small organisms present a flat relationship between species and area by which a latitudinal diversity gradient is absent or weak. Both perspectives represent one of the most debated questioning among mycologists in these days (Peay et al., 2016).

In general, there were no significant differences between microbial diversity in the different land uses along $NS_{\text{-transect}}$. Nonetheless, there were clear tendencies across the different taxa. As discussed above, *fungi* community were inclined to be more diverse in NAT than CROP ecosystems in opposition to what was plotted by *bacteria* (Figure 4-14). In addition, these pattern were more obvious towards the southmost sampling sites, *i.e.* from the towns of *Peak Hill* to *The Rock* (Site#13 to Site#23; ~500 km distance). The group of *archaea* showed a divided diversity pattern along $NS_{\text{-transect}}$ in two phases. From north to south between the towns of *Mungindi* and *Kickabil*, CROP ecosystems (Site#0 to Site#10; ~500 km distance) exhibited greater archaeal diversity than NAT ecosystems. By contrast, from the towns of *Narromine* to *Forest Hill* (Site#12 to Site#22; ~500 km distance) the above pattern changed and, in spite the high dispersion data, NAT were evidently more diverse than CROP ecosystems (Figure 4-15).

The fact that both CROP and NAT ecosystems maintained a similar tendency to increase or decrease along this transect, this may suggest that land use is a secondary factor driving microbial diversity pattern in a latitudinal large-scale perspective. As Lauber et al., (2008) findings suggested, microbial patterns changes by land use may be related to the effect of the land management practices on the edaphic properties (*e.g.* fertilization).

Microbial α-diversity along longitudinal gradient (WE-transect)

The main microbial diversity pattern found along the longitudinal gradient was a marked fluctuation in *fungi* and *archaea* groups, which contrasted with a slightly steady pattern observed in the *bacteria* (Figure 4-13 to Figure 4-15). In general terms, fungal, archaeal and bacterial diversities were higher in the west of NSW and lower toward the coastal zone (*i.e*. *Coffs Harbour*). In *archaea*, diversity reduction occurred gradually from west to east, and also showed evident differences by land use. For example, this group reduced diversity values from PD ~2.8 to 0 (but PD= ~1.7 in CROP sites) (Figure 4-15). In *fungi*, the diversity pattern oscillated in two phases. For example, from west to east in natural environments *Chao1* values oscillated from ~ 832 to ~ 651, then increased to ~ 956 before decreasing to approximately 566 on the east coast (Figure 4-14). In contrast to both *fungi* and *archaea* patterns, the diversity of *bacteria* did not show significant fluctuations from west to east: over approximately 650 km along the $WE_{\text{-transect}}$ PD values ranged between ~345 to ~336 (Figure 4-13). Only near *Narrabri* (Site#35) did bacterial diversity show any decrease, reaching PD ~ 291 in the *Coffs Harbour* area.

Although similar tendencies were observed on richness and evenness metrics, the diversity patterns described above were slightly less apparent when abundance data was included using *H'* and *Simpson* metrics. This may indicate that rare and common species could be following different patterns.

The effect of land use on the microbial community along this $WE_{-transect}$ was significant in the westernmost sampling sites, *i.e.* between the towns of *Wanaaring* and *Bourke* (Site#48 to Site#44). However, it must be considered that the so-called cropping ecosystems in these locations were not necessarily cropping areas. Indeed, they were mainly uncovered zones, as agroecological conditions in these areas are not generally conducive to rainfed agriculture (see Chapter 2). In these western zones, there were greater diversities of *fungi* and *bacteria* in natural/undisturbed than cropping/disturbed ecosystems (Figure 4-13 and Figure 4-14). Respectively, diversity values in NAT and CROP ecosystems oscillated ~331>PD>~275 in *bacteria*; and; ~891>Chao1>~624 in *fungi*. At an enormously lower level, *archaea* ranged ~2.4<PD<~3.2 in NAT and CROP, respectively (Figure 4-15).

Toward the easternmost sampling sites, however, across the three kingdoms these patterns were not clear. We found sometimes CROP ecosystems showing greater diversity levels than NAT ecosystems. This result was mainly observed in the evenness values by taking into account abundant species ((Figure 4-13 to Figure 4-15). Surprisingly, these first impressions demonstrated a completely unexpected decrease of microbial diversity towards the most vegetated and humid areas of the eastern NSW. Moreover, since bacterial diversity did not show a strong fluctuation pattern, they did not seem to be strongly affected by rainfall, temperature or altitudinal factors which are characteristic gradients of $WE_{-transect}$. As described in Chapter 2, $WE_{-transect}$ was designed completely orthogonal to NSW rainfall pattern fluctuating from ~170 to 3,200 mm/year.

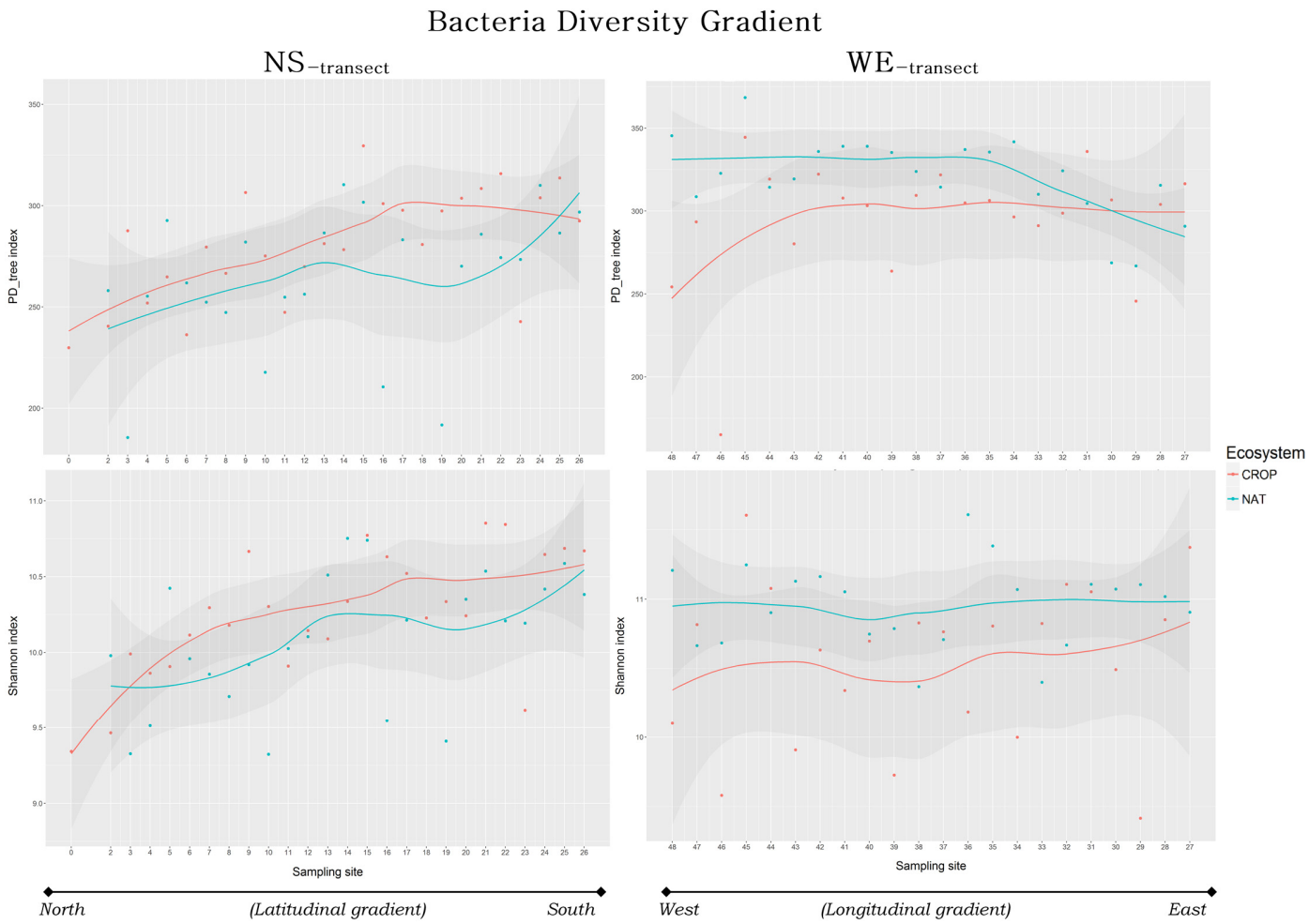


Figure 4−13. *Bacteria* diversity gradient along NS−transect and WE−transect. The fitted lines show a regression model (Wickham, 2009) upon the mean diversity values from 10 iterations obtained in QIIME (Caporaso et al., 2010b).

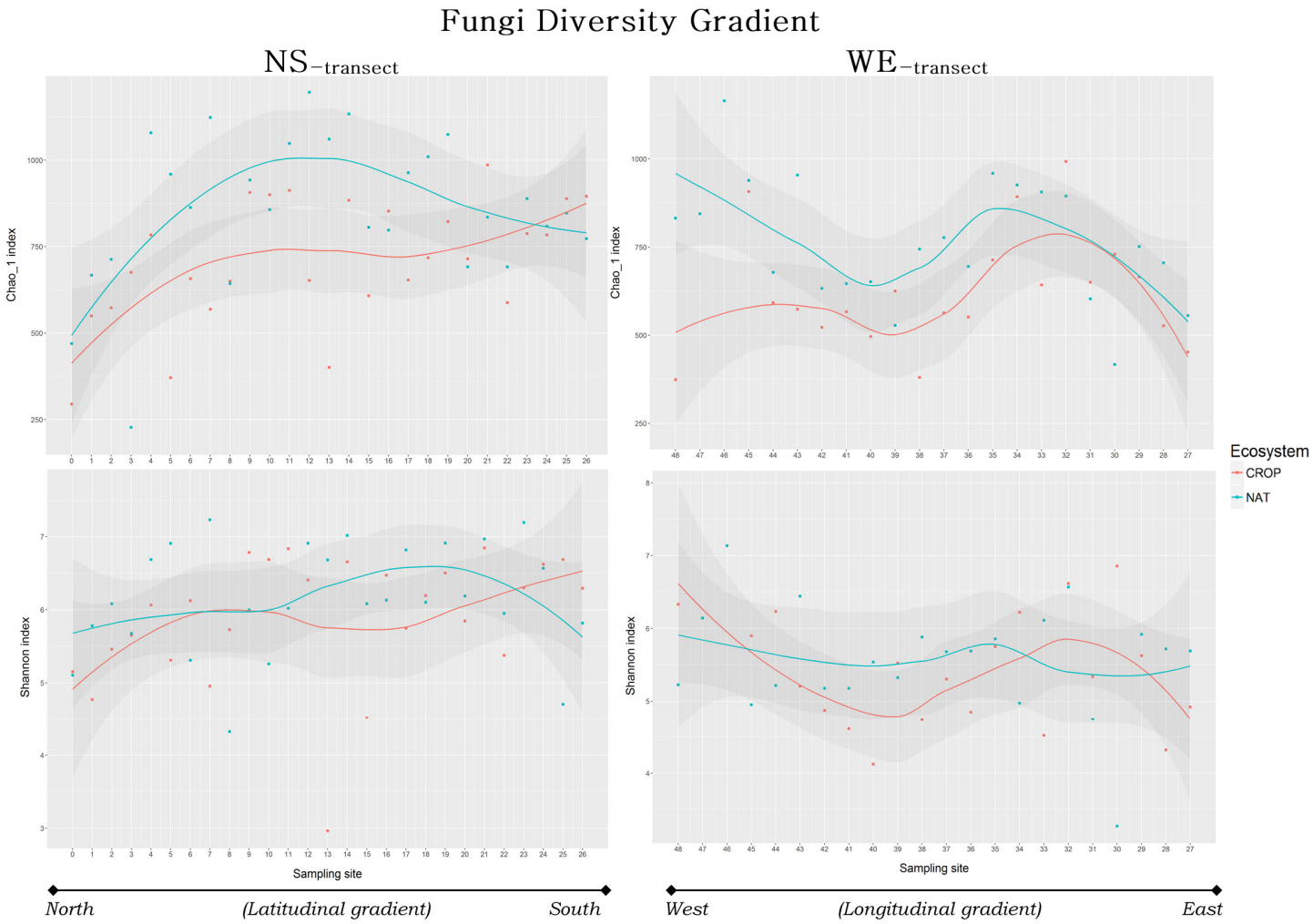


Figure 4−14. *Fungi* diversity gradient along NS−transect and WE−transect. The fitted lines show a regression model (Wickham, 2009) upon the mean diversity values from 10 iterations obtained in QIIME (Caporaso et al., 2010b).

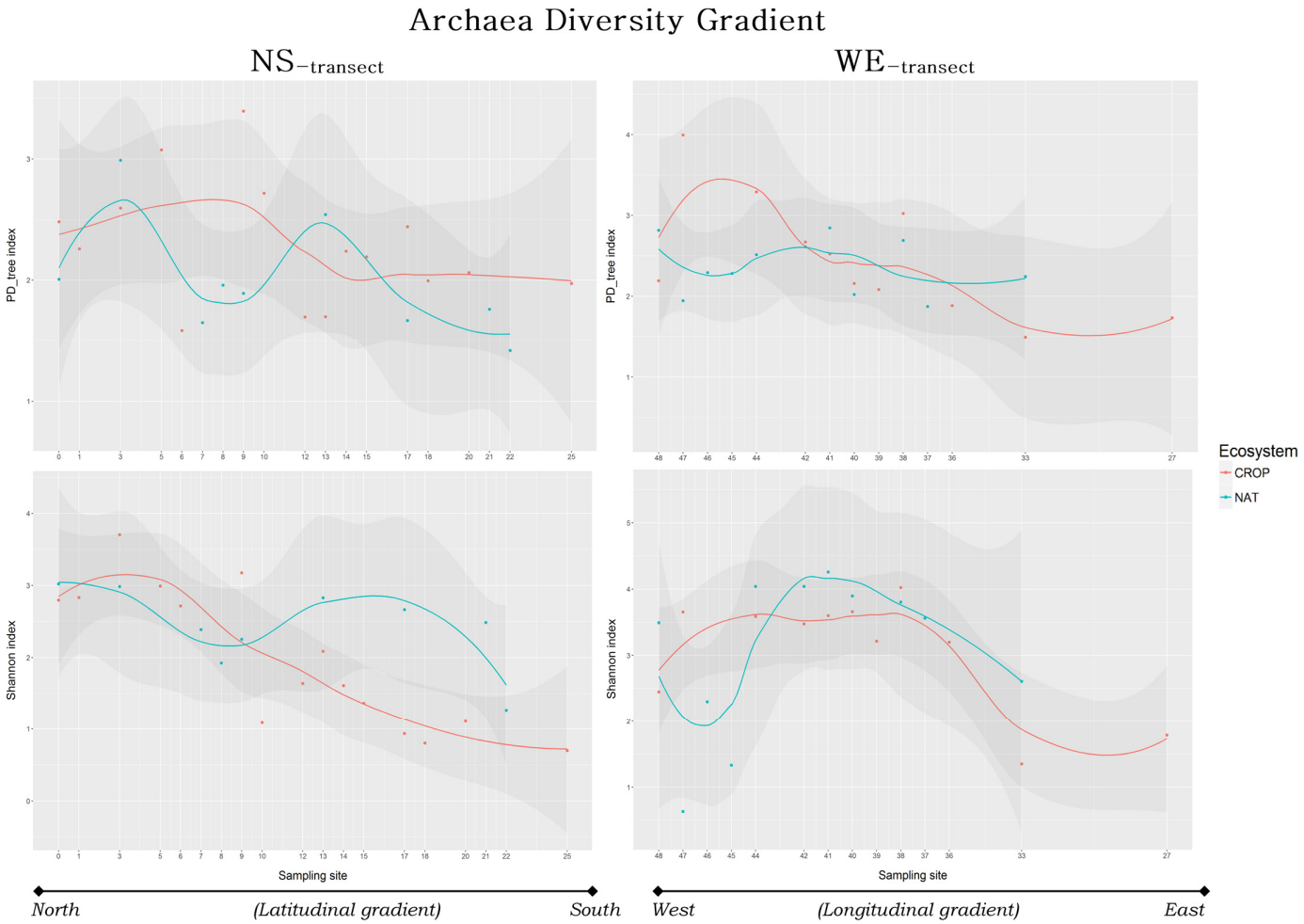


Figure 4−15. *Archaea* diversity gradient along NS−transect and WE−transect. The fitted lines show a regression model (Wickham, 2009) upon the mean diversity values from 10 iterations obtained in QIIME (Caporaso et al., 2010b)

E. MICROBIAL DIVERSITY AND ABIOTIC PARAMETERS RELATIONSHIP

Soil microbial structural patterns (richness, evenness, composition) were assessed in relation to environmental gradients characterising NSW. In this direction, and assuming the multifaceted network between biodiversity and pedodiversity, we decide to explore and probe this relationship in order from more simple to complex analysis. Thus, the microbial diversity and abundance data, soil physicochemical properties (*e.g.* pH, TC, TN, etc.), and other environmental covariates (*e.g.* rainfall, temperature, etc.) information have been analysed from simple linear associations (e.*g.* linear correlations) to more complex multidimensional exploratory analyses (*e.g.* PCAs, bootstrap regression modelling, and mapping).

Soil microbial patterns and physicochemical attributes direct relationship

In general, we detected more cases of high significant correlations between microbial diversity and soil attributes than with other determinant environmental factors. Indeed, *fungi* diversity and the abundance of *archaea* were almost exclusively affected by soil properties (Table 5-5 and Table 5-6). For instance, fungal diversity was highly correlated with soil physicochemical properties ($r=-0.48$; $p\leq0.05$) rather than with other primary environmental covariates, e.g. land surface temperature ($r=-0.26$; $p\leq0.05$). In order of magnitude, the soil variables highly correlated with fungi diversity were aggregation (SI), exchangeable Ca, clay and soil pH ($r=-0.48$, -0.43, -0.41 and -0.41; $p\leq0.05$) (Table 5-6). Moreover, SI and pH together with TC were affected significantly fungal abundances ($r=-0.34$, -0.28, and 0.29; $p\leq0.05$) (Table 5-5). Archaeal diversity

was strongly correlated with attributes of land surface temperature, elevation and soil pH (r=0.69, -0.67 and 0.66; p≤0.05), which were closely followed by slope, aggregation (SI), exchangeable Mg and ECEC (r=-0.64; 0.60, 0.59 and 0.57; p≤0.05). Even though the abundances of these prokaryotes was the less correlated with both soil and environmental gradients, the few significant correlations were observed with wetness index (r=0.27; p≤0.05) and other measured soil properties (Table 5-5). In fact, the most abundant phyla *Thaumarchaeota* showed only a significant negative correlation with exchangeable Na (r=-0.23; p≤0.05). By themselves, the least abundant phyla, besides of wetness index (r=0.27; p≤0.05), has shown important associations with exchangeable Na, NH_4-N, exchangeable Al, exchangeable K and T-Hor (r=0.47, 0.39, 0.31, -0.26 and 0.25; p≤0.05).

Bacterial diversity was observed highly related with exchangeable Al, Band 5 (Landsat7), soil pH and K (r=-0.43, 0.41, 0.40 and 0.39, respectively; p≤0.05). More than other taxa, the relative abundance of *bacteria* was greatly influenced by both environmental and soil attributes. After elevation (r=0.57; p≤0.05), the highest relationships with the relative abundance of bacteria phyla were found with land surface temperature, soil pH and exchangeable Ca (r=-0.52, -0.51 and -0.51, respectively; p≤0.05).

So far, we have found soils attributes notably influencing diversity and abundances of microorganisms. These relationships were as important as other crucial environmental variables (*e.g.* elevation), and in some cases even higher. For example, the strongest correlation of the abundant *Actinobacteria* was equally strong and significant with land surface temperature and soil pH (r=0.45; p≤0.05). This same results were documented by Valverde et al.(2012) in hot

springs in Zambia, China, New Zealand and Kenya.

Nonetheless, the degree and direction at which soil properties and microbial communities showed to be related to each other has been variable depending on both microbial taxa (*e.g.* kingdom, phyla) and type of diversity analysed (*e.g.* richness, evenness or phylogeny) (Table 5-6). For instance, at kingdom level, soil pH was one of the most consistent soil attributes affecting diversities across the three kingdoms and, it has been particularly high in correlation with archaeal diversity (*e.g.* $0.48 \leq r \leq 0.66$: $p \leq 0.05$). However, soil pH did not impact significantly abundances of *archaea* ($-0.15 \leq r \leq 0.22$; $p \geq 0.05$) but did affect *fungi* and *bacteria*. Similarly, at phylum level, and as documented by other investigations (Lauber et al., 2009), we observed the strong but opposite responses of *Actinobacteria* (positively related) and *Acidobacteria* (negatively) to soil pH gradients. In the same way, exchangeable Na, for example, was negatively correlated with the abundant *Thaumarchaeota* but positively correlated with the relative abundance of the rare *Thermoplasmata,* the strongest relationship found in kingdom *archaea* ($r=0.47$, $p \leq 0.05$) (Table 5-5).

We also observed that abiotic parameters affected microbial diversity differently. For instance, clay content of soils was strongly related with both richness and evenness of *archaea* ($r=0.28$; 0.48 and 0.53, respectively for Chao1, H' and 1-D; $p \leq 0.05$) but it did show an important impact on their phylogenetic diversity ($r=0.1$ for PD, $p \geq 0.05$) (Table 5-6). In addition, clay and *Archaean* abundances correlations was weak and not significant ($p \geq 0.05$) (Table 5-5).

Both microbial abundances/diversity and soil attributes linear (Figure 4-16 to Figure 4-21) and

quadratic (Figure 4-22 to Figure 4-24) relationships were exemplified along the extent of some soil physical and chemical gradients found in our study area.

From these analyses, we concluded that both linear and non-linear associations between microbial diversity/abundance and soil properties were quite evident but it was difficult to generalize the main soil properties controlling microbial patterns. Indeed, these microbial patterns apart of being variable among microbial kingdoms and taxa (e.g. phyla, genera) were also variable when responding to the set of soil properties analysed. For example, in general terms, there was a high influence of soil properties controlling archaeal diversity but they scarcely were seen affecting archaeal abundance. In contrast, bacterial abundance resulted to be more commonly and significantly affected by soil attributes than bacterial diversity. In fungi, both diversity and abundance were similarly affected by soil traits. On top of that, the effect of these soil attributes were variable across the different microbial phyla. For example, the dominant phylum *Ascomycota* was not significantly influenced by soil chemical attributes (e.g. soil pH) but by physical attributes, instead; and *Basidiomycota* was strongly affected by most of the soil physiochemical properties analysed.

Similar conclusions have been found in other investigations (Constancias et al., 2015; de Gannes et al., 2015; Garbeva et al., 2004; Ranjard et al., 2013). For instance, de Gannes et al., (2015) after their work on analysing *bacteria*, *archaea* and *fungi* communities across different series of tropical soils in Trinidad and Tobago indicated that '*microbiomes in each soil acquired unique identities*'. From their findings, they also concluded that no a single soil attributes such as soil pH can explain microbial community patterns by itself but other soil factors should be included.

The fact that 'microbiomes' acquired unique identities in each soil might imply that, for example, the diversity of the microbial communities should not be significantly variable within similar soil entities, i.e. low pedodiversity. For example, this notion agreed with Rampelotto et al., (2013) investigation in which they could not find significant bacterial diversity changes along the same soil type. Lastly, these author gave more relevance to bacterial composition dissimilarities rather than diversity patterns and, they concluded that *land use* was the main factor controlling these dissimilarities. In this sense, we can add that any effect of land use is merely a combined modification of multiple soil properties i.e. an alteration of the microbiome. This notion was supported by Lauber et al., (2008) who had argued that the influence of land use-types on microbial communities comes directly from their impact on the edaphic properties; yet edaphic properties are always involved defining the soil composition.

In our research, we argued that all of these perspectives pointed out the importance of soil properties, and that both diversity and abundance of microbes might be related with certain groups of soil properties commonly associated to determined soil entities (*e.g.* soil classes, soil types). We explored this argument by plotting the distribution of soil microbial richness (Chao1) across the six soil orders (*i.e.* at the highest level of the soil Australian taxonomy) encompassed by the 49 sampling sites in our study area (Figure 4-25) (see Australian Soil Classification distribution in NSW in Chapter 2).

Despite the limited data, we observed the microbial structural patterns of the three microbial kingdoms in *Vertosols*, *Sodosols*, *Chromosols* and *Kandosols*. From this analysis, we observed that microorganisms tended to be more diverse and abundant in both *Vertosols* and *Sodosols* than in

the other soil types encompassed in our study area. *Bacteria* and *archaea* were highly abundant and diverse in *Vertosols*. *Fungi*, instead, ranged higher abundance and diversity in *Sodosols* but lower values in *Vertosols*. Immediately, these soil types were followed *Tenosols* and *Rudosols* in which, the lower diversity and abundance of both *bacteria* and *fungi* was evident as such the complete absence of *archaea*. From our point of view, it is important to highlight such tendencies despite the fact that all of these soil orders are predominantly classified based on the attributes of B horizons; yet our microbial data comes from A horizons. Even though we found similar microbial diversity and abundance following similar patterns, the diversity were slightly less pronounced on the basis of the Shannon diversity (H') estimator.

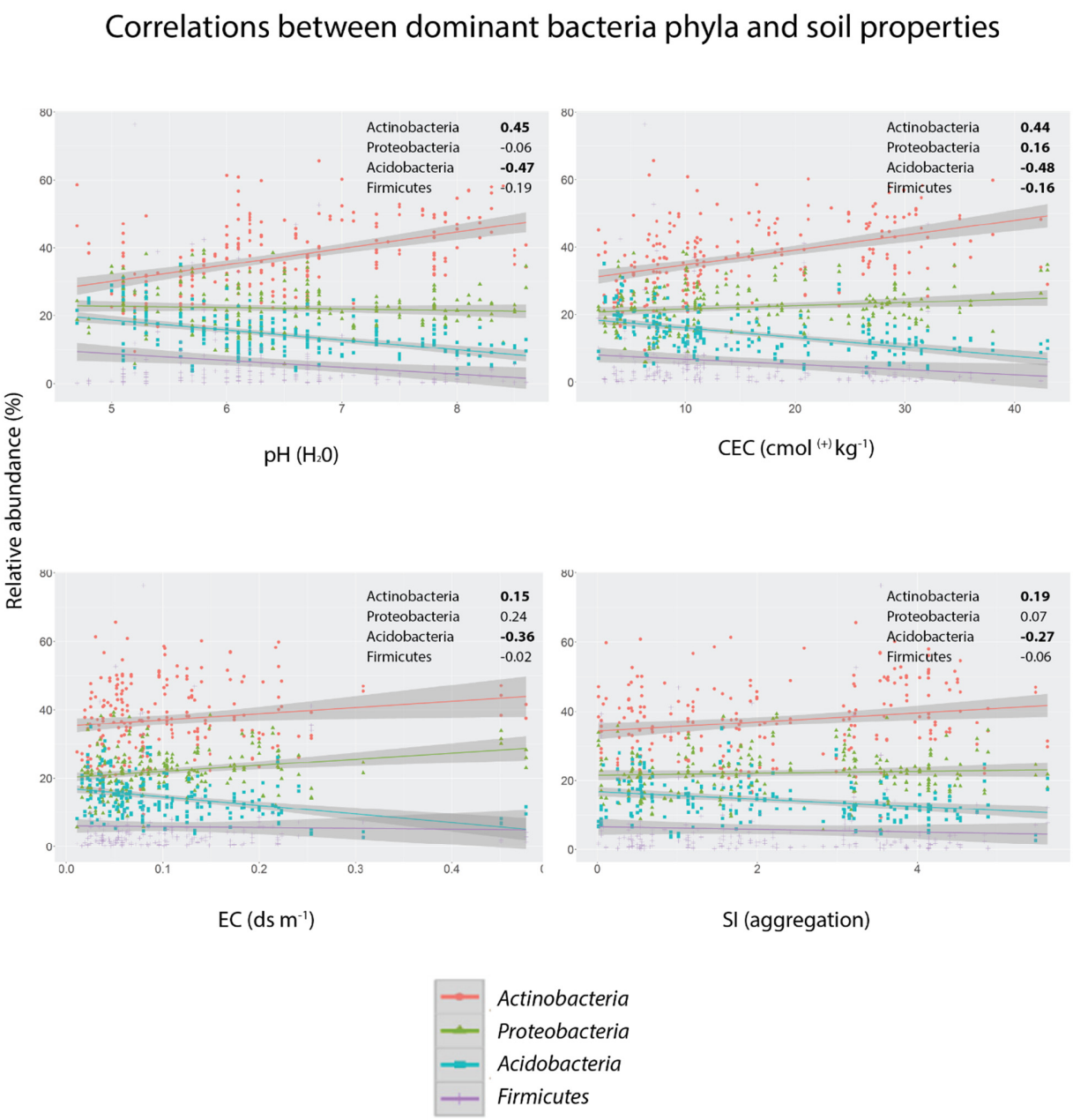


Figure 4−16. Correlations between the relative abundance of dominant bacteria and soil physicochemical properties.

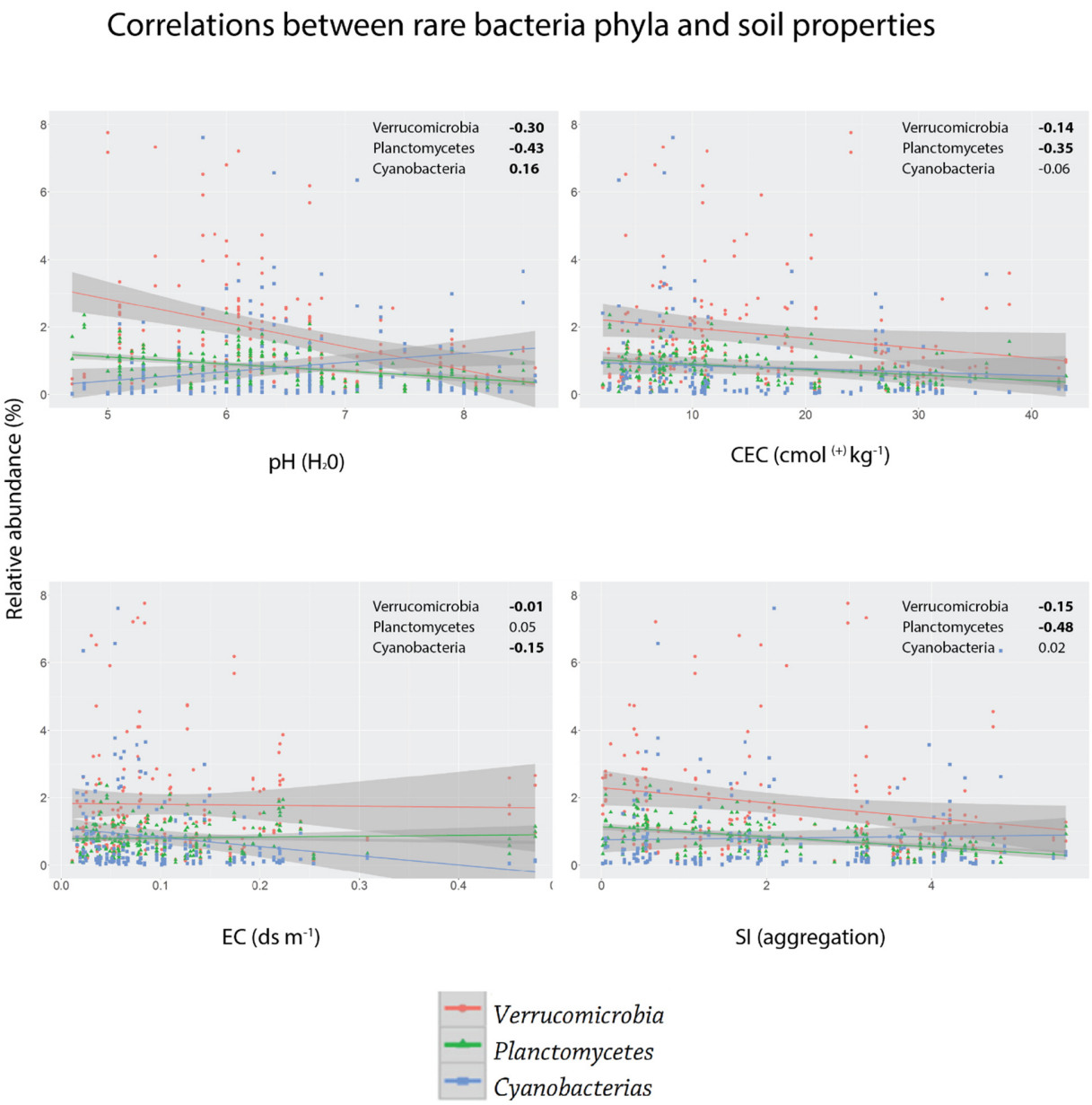


Figure 4-17. Correlations between the relative abundance of rare bacteria and soil physicochemical properties.

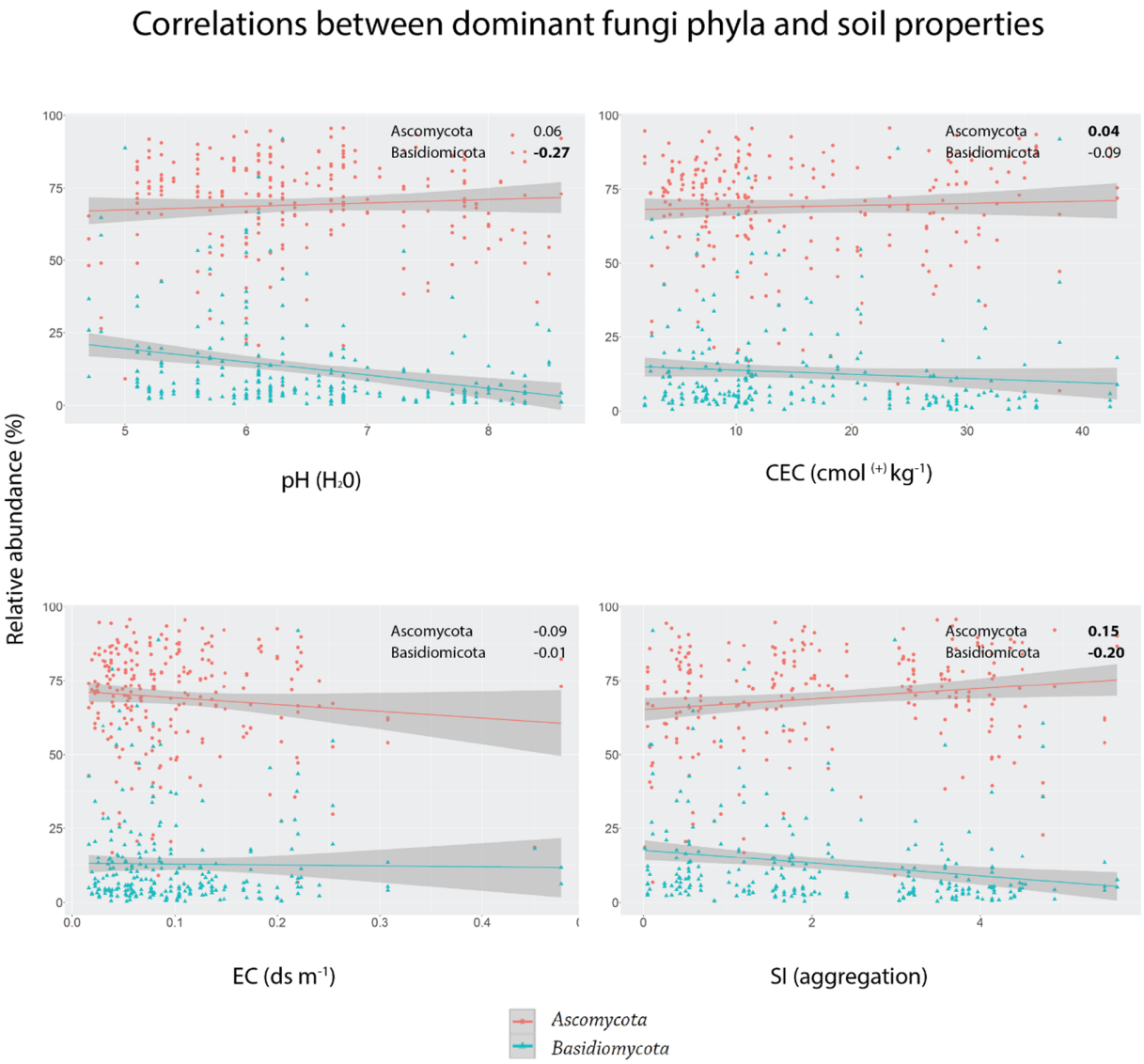


Figure 4−18. Correlations between the relative abundance of dominant fungi and soil physicochemical properties.

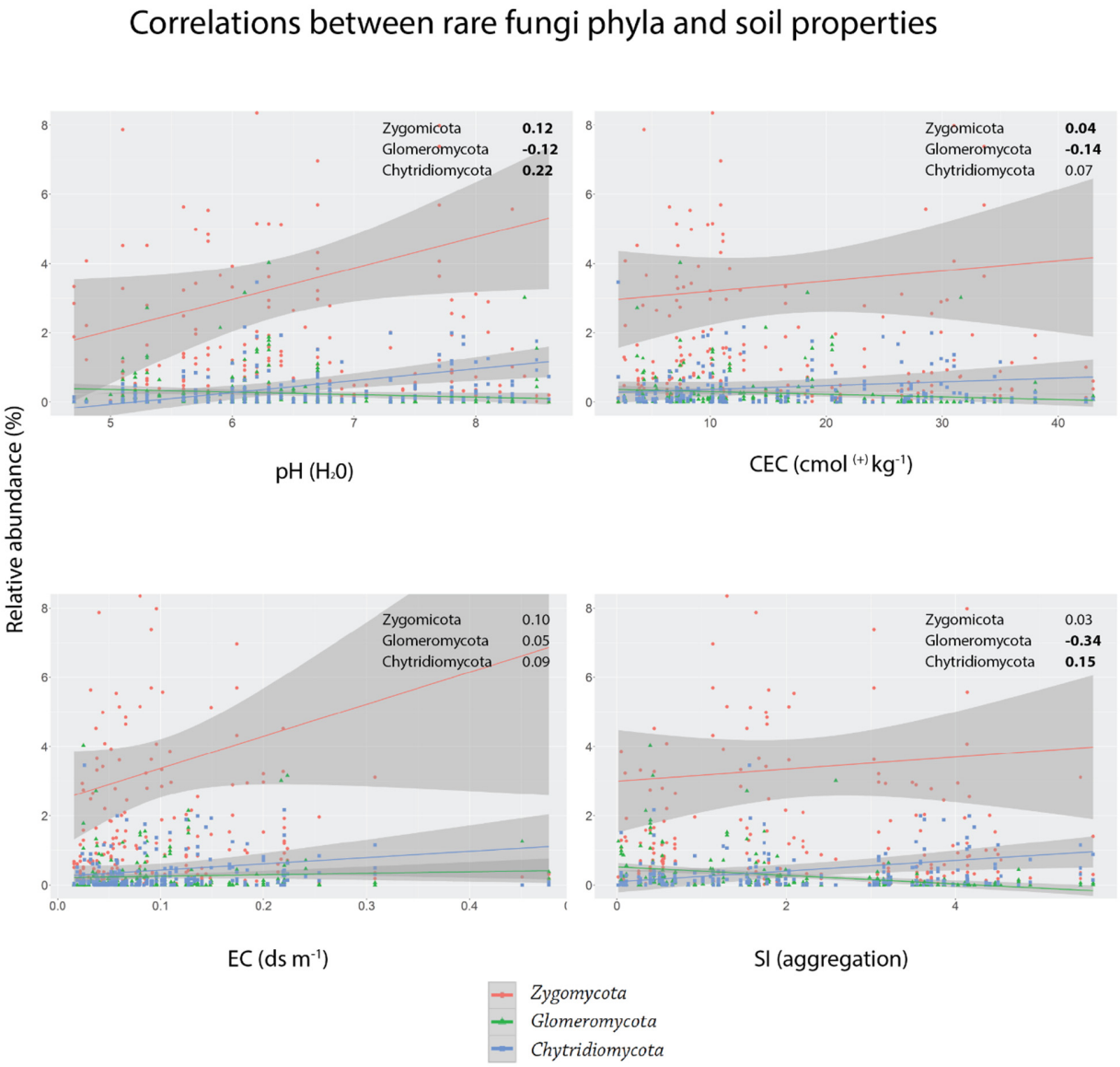


Figure 4-19. Correlations between the relative abundance of rare fungi and soil physicochemical properties.

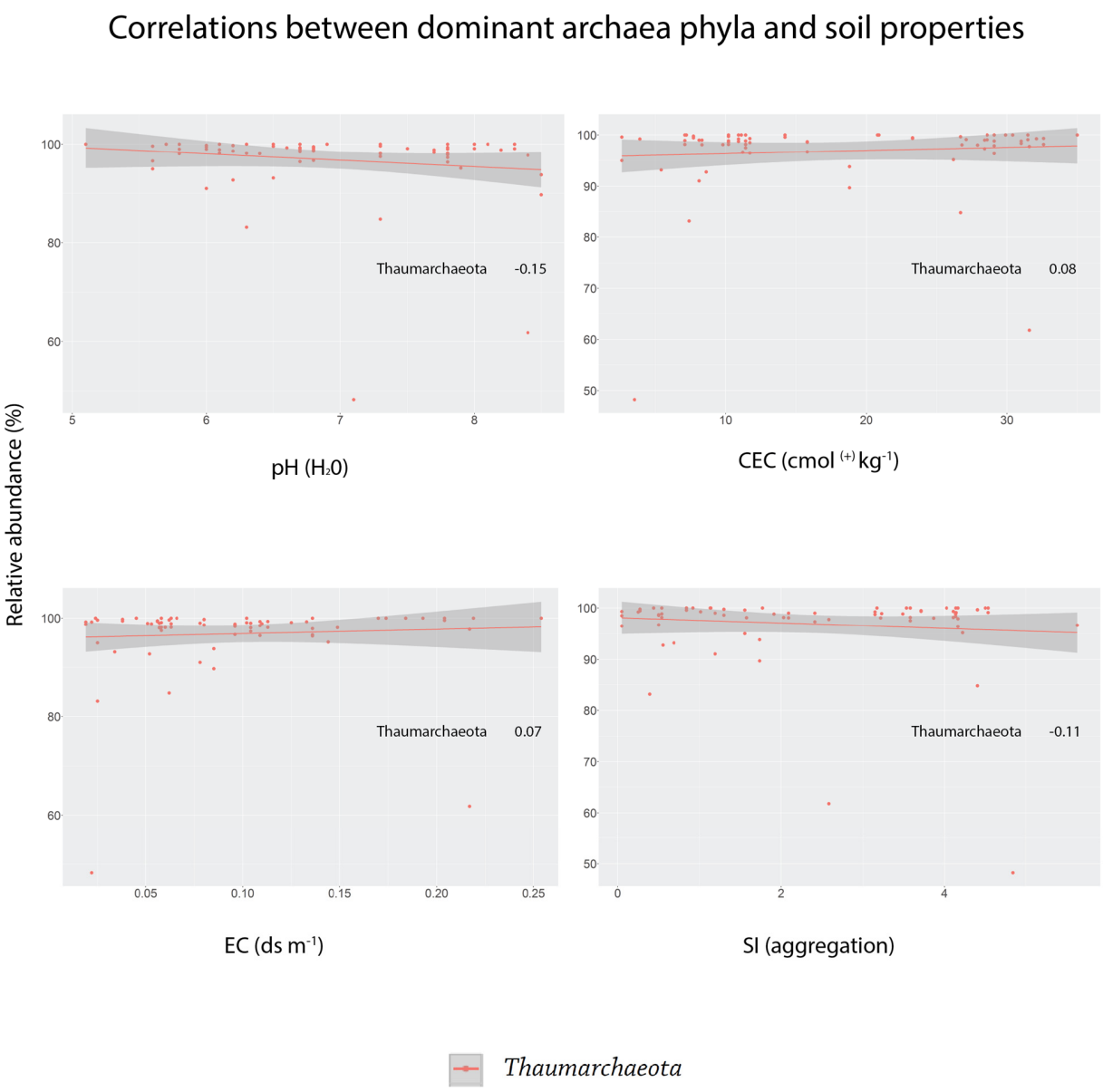


Figure 4−20. Correlations between the relative abundance of dominant archaea and soil physicochemical properties.

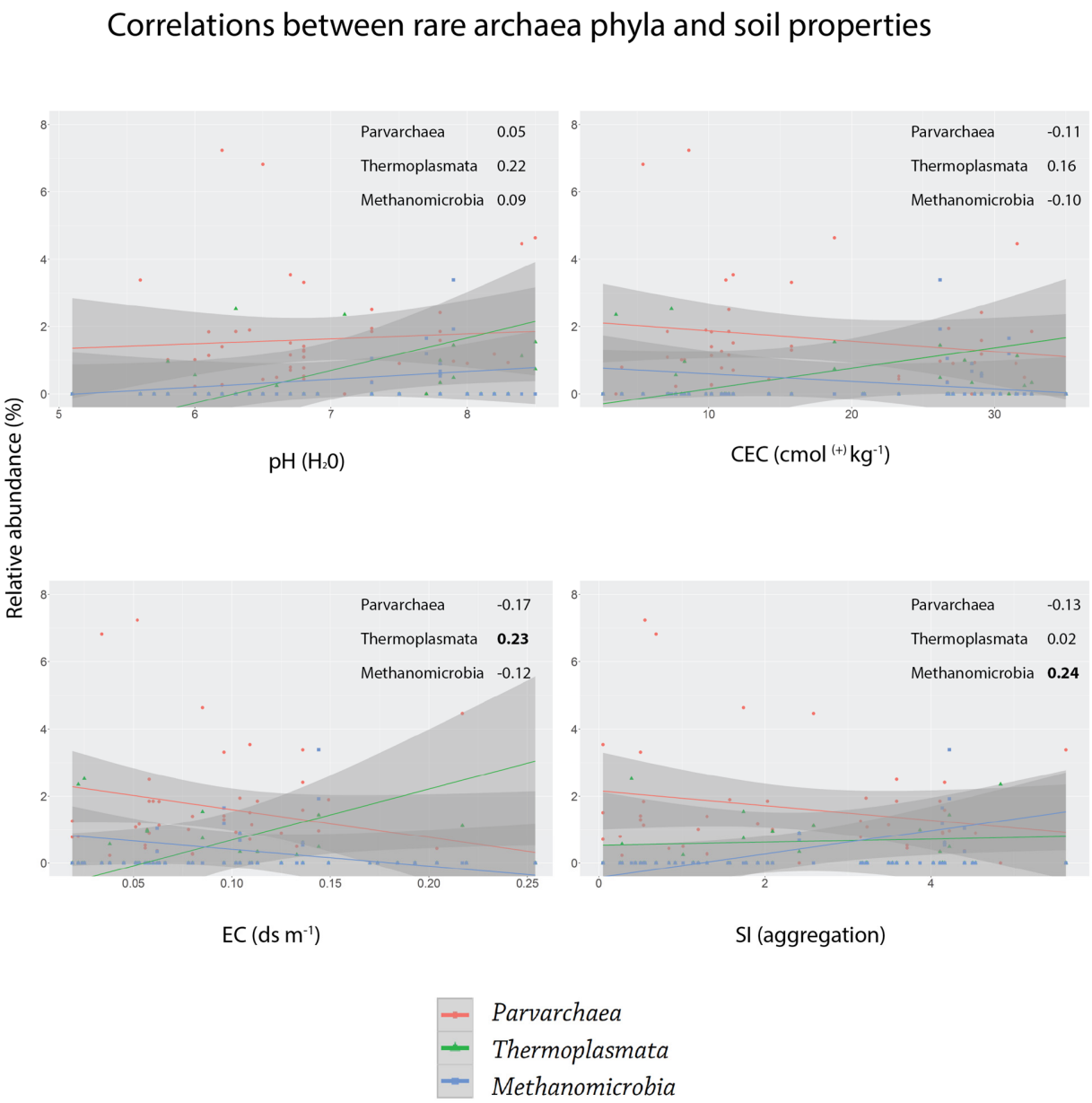


Figure 4-21. Correlations between the relative abundance of rare archaea and soil physicochemical properties.

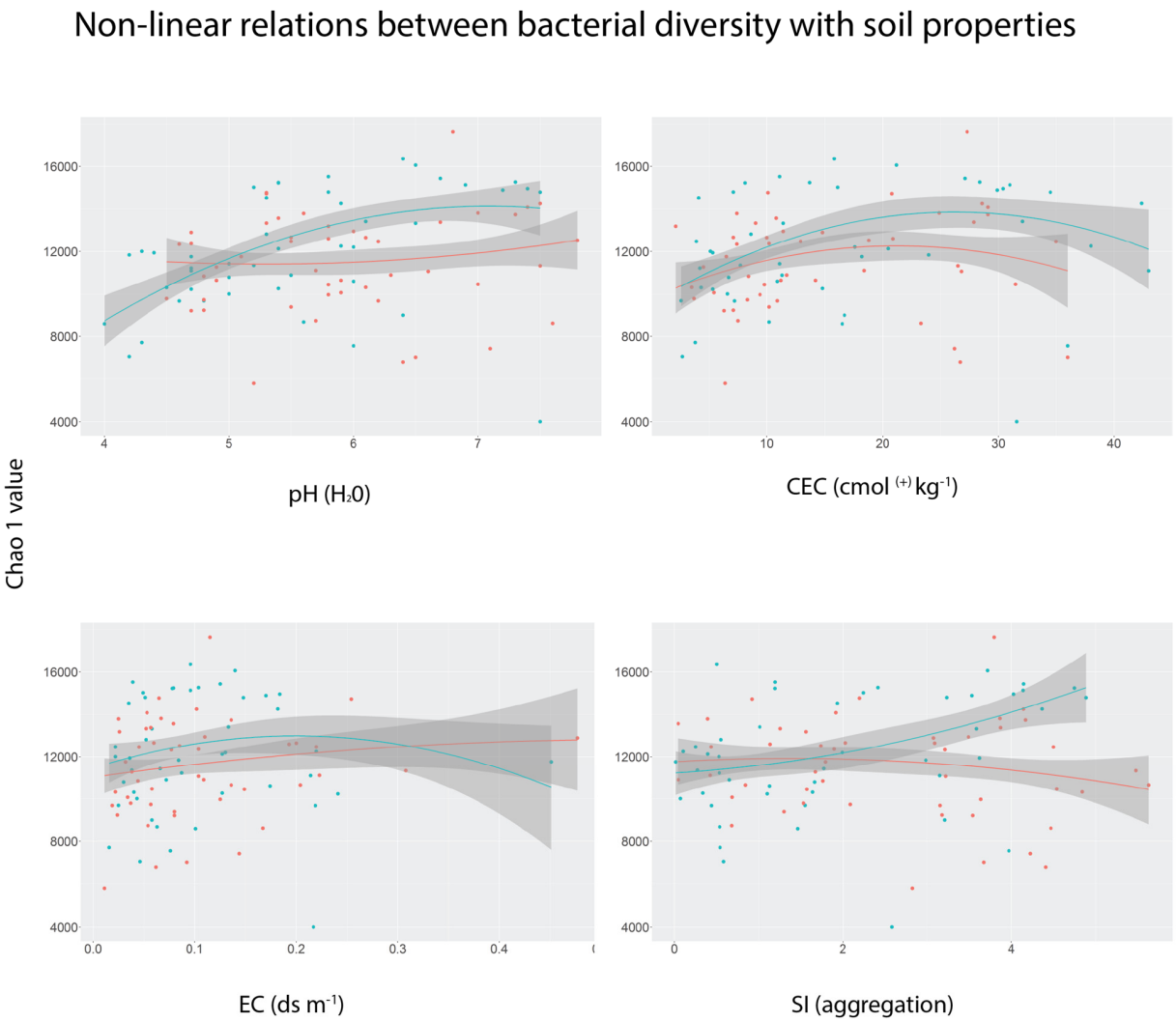


Figure 4−22. Quadratic associations between bacterial diversity and soil physicochemical properties.

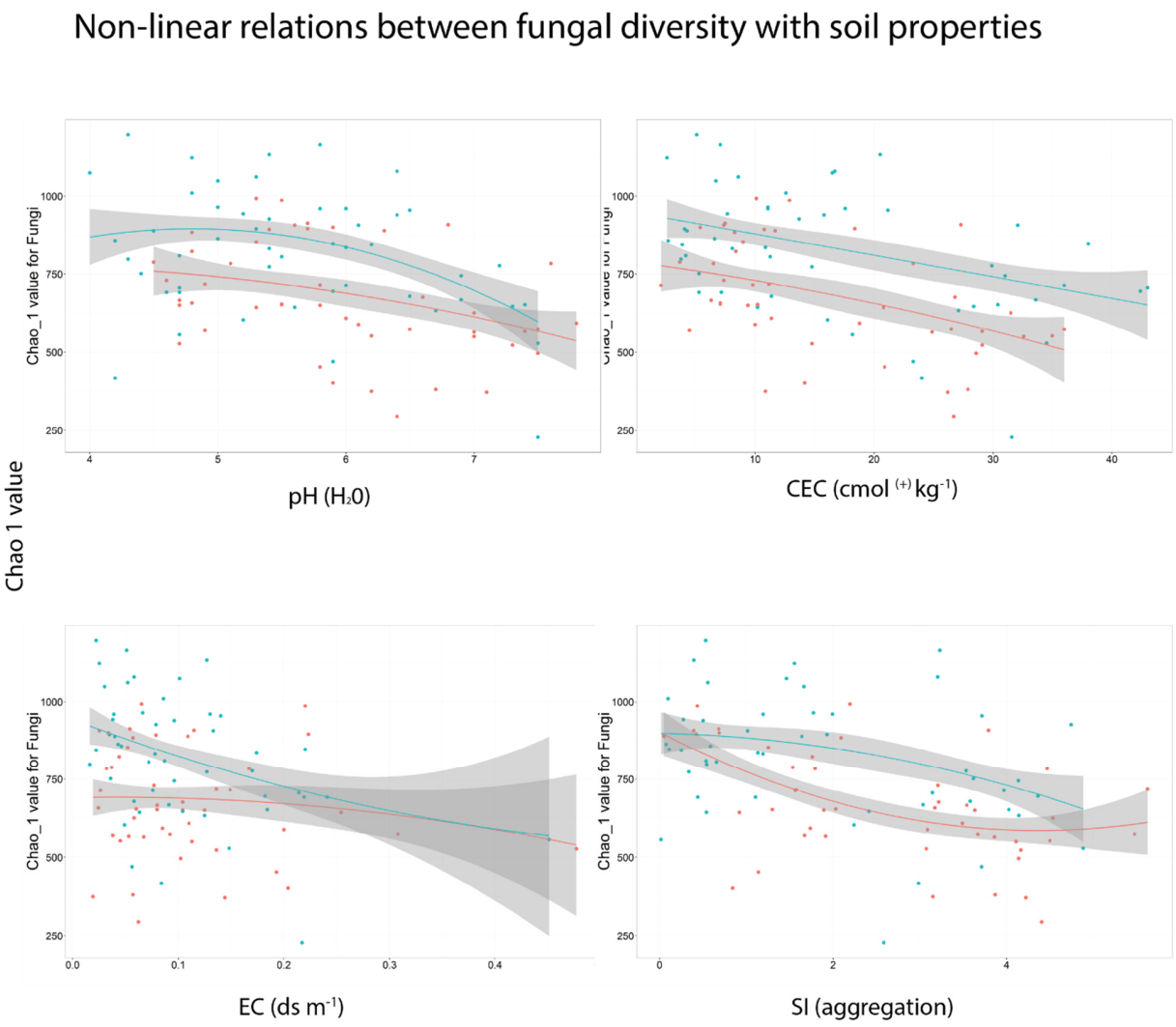


Figure 4-23. Quadratic associations between fungal diversity and soil physicochemical properties.

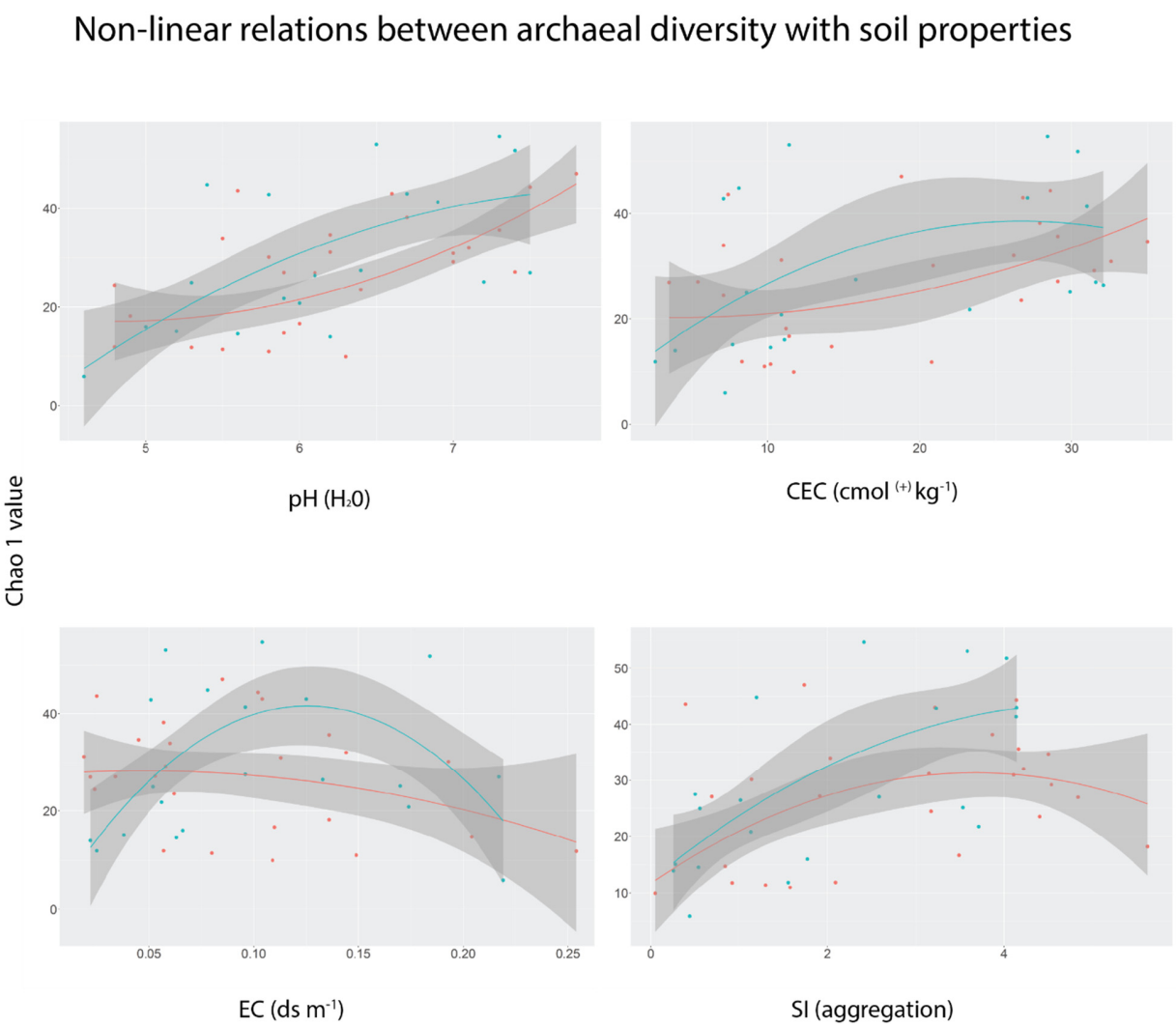


Figure 4−24. Quadratic associations between archaeal diversity and soil physicochemical properties.

Soil microbial distribution across soil orders sampled from our study area

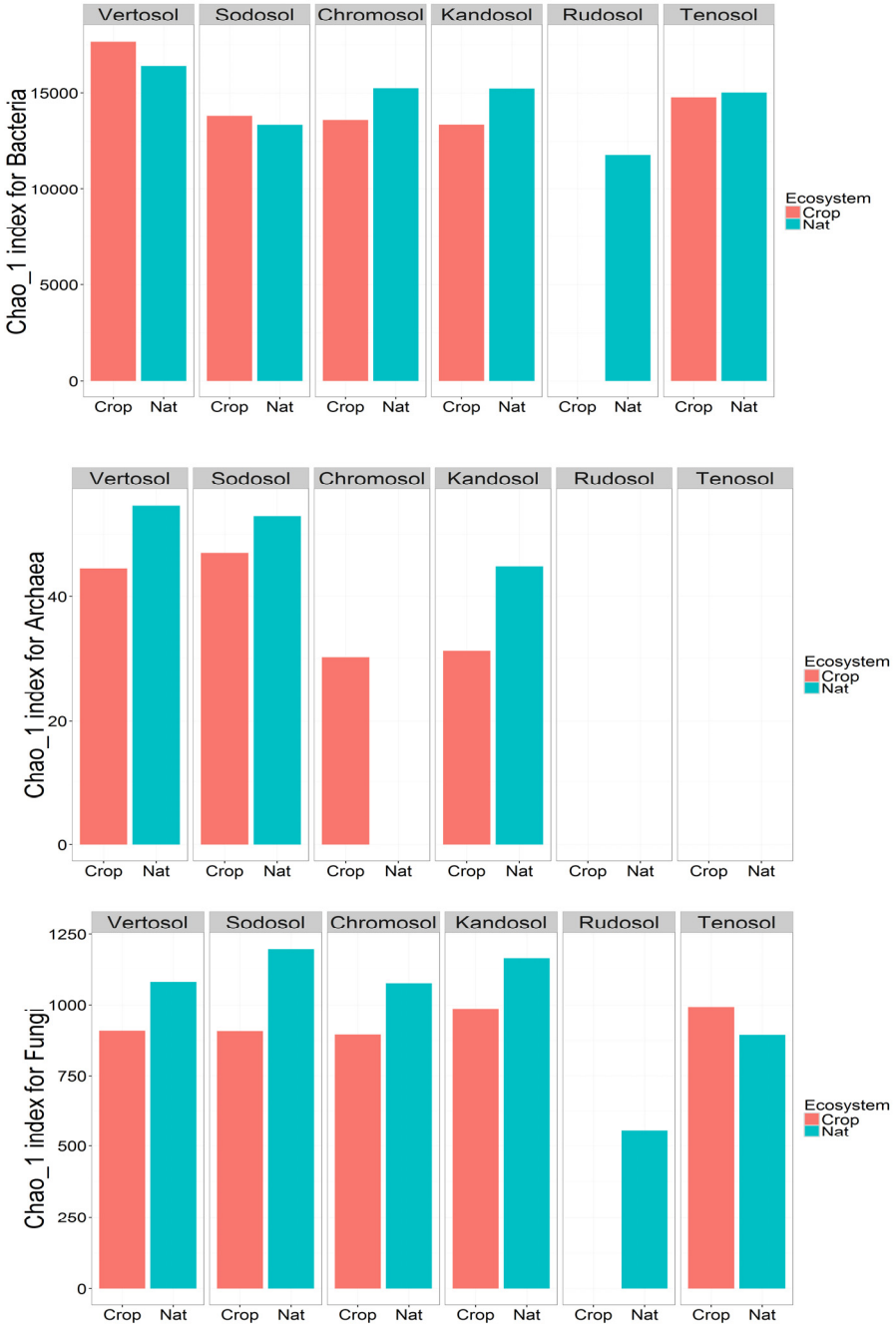


Figure 4−25. Microbial richness distributed across the different NSW soil orders according to the Australian Soil Classification Systems (ASC).

Multivariate correlations between microbial diversity and soil attributes:

By assuming the complex association between microbial communities and soil characteristics, we evaluated them in a multivariate analysis using principal components (PCA). This analysis was made in R (R Team et al., 2013) using the "prcomp" package using prior scaling and centering the data.

In the PCA analysis, soil microbial diversity was highly explained by soil physicochemical attributes (variance explained≥0.54) and slightly more by other environmental covariates (variance explained ≥0.6) – although these plots are shown together in Figure 4-26 to Figure 4-31, they were also analysed separated into soil and environmental covariates. Despite this result, the strongest relations for all kingdoms were found with soil properties. A general pattern we observed was that in all the PCA plots both soil attributes and environmental covariates tended to be grouped representing certain either soil entities or environmental ecoregions. For example, clay, Ex_Na, pH and SI were seen always grouped as they are commonly found defining *Vertosols* or *Kandosols*. This same pattern was seen with certain set of other variables representing zones of our transects. For example, TC, TN, NDWI, NDVI tended to be clustered, probably, representing forested zones of the eastern zone (see Chapter 2, ecoregions). What is more, the orthogonality of our transects exposed the most distinctive environmental attributes of them facing each other when both transects were analysed separately. For example, on the $NS_{transect}$ PCAs showed attributes of *Vertosols* (located in the north) facing attributes such as slope,

elevation and NDVI characterising the southern part. Similarly, $WE_{-transect}$ PCAs showed temperature always facing precipitation, etc. Therefore, in this context we evaluated microbial diversity indices which used to display strong relations with these clusters of attributes.

From the PCAs we found diversities of *archaea* and *fungi* were better explained by soil attributes than diversity of *bacteria (*variances explained= 0.58; 0.55 and 0.54, respectively*)*. In a latitudinal gradient ($NS_{-transect}$) *bacteria* showed strongest relations with TC, TN, NO_3, B2-B5-B4 (Landsat7) but weak relations with Clay, Ex_Na and SI (Figure 4-27). In our longitudinal gradient ($WE_{-transect}$) when precipitation and temperature tended to cancel each other, *bacteria* were mostly related with ECEC and most of the exchangeable cations. Remarkably, the strongest pattern seen in bacteria was a positive relation of PD diversity with clay (and their associated SI and Ex_Na) and a negative relation with *T-Hor* (depth of the first horizon). Bacterial Shannon diversity was shown to be positively related to P and EC (Figure 4-27).

A common pattern found with *fungi* was a robust positive relation with T-Hor, for which, the diversity might be influenced by soil depth. Along $NS_{-transect}$ there was a high positive association with Ex_Al and negative with soil pH. They also showed negative relation with temperature, clay, ECEC, SI and other attributes characterising the northern part of this transect (Figure 4-28). In $WE_{-transect}$, *fungi* diversity was highly positive related with B1-B2 (Landsat7; vegetation) and once again strongly negative related with ECEC and soil exchangeable cations Ca, K and Mg (Figure 4-29).

The diversity of *archaea* was equally related with soil properties as with environmental attributes

(both variance explained=0.6). This kingdom's Shannon diversity in both transects showed great positive association with clay, Ex_Na, pH and SI and other exchangeable cations. However, their phylogenetic diversification showed to be closer and positive in relation with temperature. This differentiation was more evident in $WE_{\text{-transect}}$ were temperature gradient is broader. This was the only kingdom displaying different patterns between those two diversity metrics. This results might suggest that there are different components controlling archaeal abundances and phylogenetic diversification. In addition, along the latitudinal gradient ($NS_{\text{-transect}}$) we found diversity of *archaea* in a significant and negative relation with slope, TC, TN, B4 and B5 (Landsat7 bands usually associated with shoreline's vegetation and discrimination between soil and vegetation moisture). Whereas along $WE_{\text{-transect}}$ gradient archaeal diversity was negatively related with attributes characterising coastal areas, *i.e.* precipitation, NDWI, wetness.

All the tendencies and significant relations obtained from PCAs analyses were quite consistent with the patterns found in the lineal and quadratics associations described earlier (Figure 4-24). The advantage of PCAs analyses was the exercise of grouping distinctive attributes characterising our study area. Furthermore, these result were also quite consistent with the pattern we found when characterising the diversity gradients changing along the transects (Figure 4-13 to Figure 4-15). So far the evidence does not suggest any singular driver of soil microbial diversity neither within the soil attributes or within the other covariates. Instead, they showed a connection with a set of them representing determined ecosystems. However, we also observed that the set of variables representing strong properties related to *Vertosols* in the north of NSW were one of the most substantial aspects affecting soil microbes. Additionally, diversity and abundance of microorganisms not necessarily can be controlled by same set of attributes.

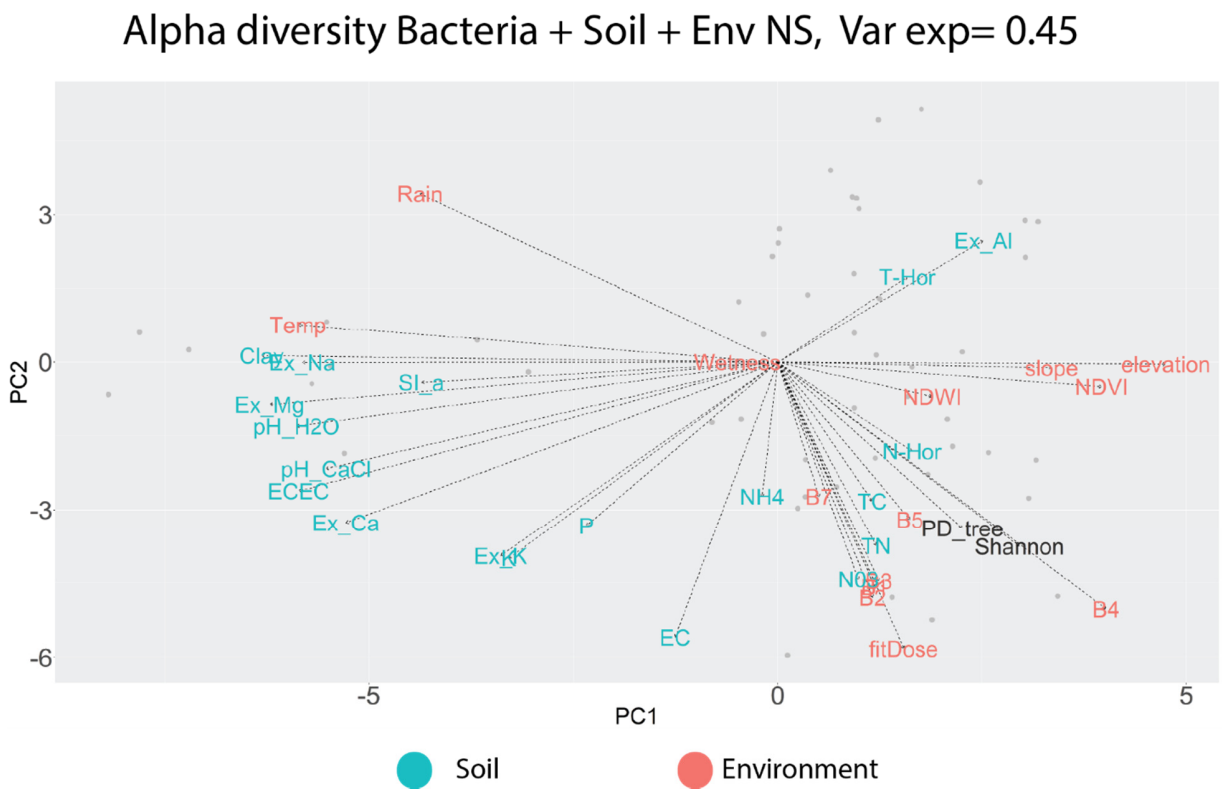


Figure 4-26. Principal component analysis of the bacteria α-diversity with soil physicochemical attributes and other environmental variables for NS-transect.

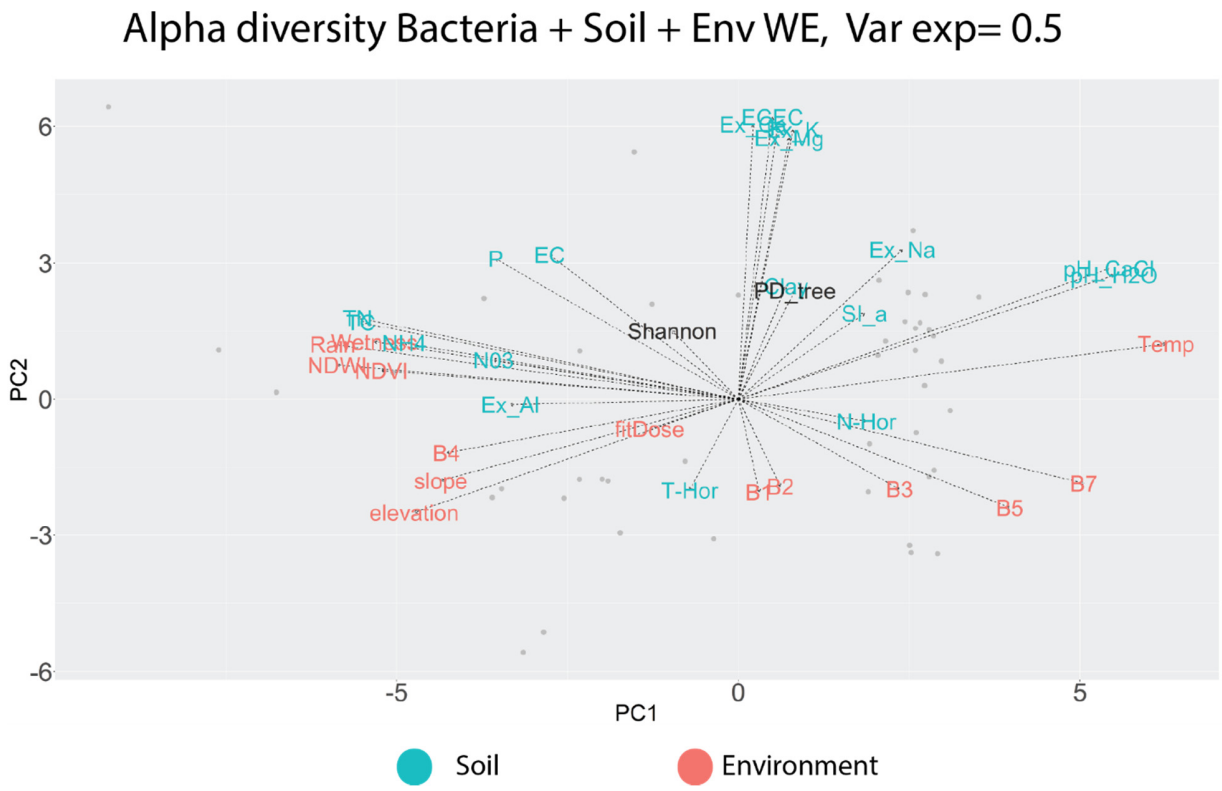


Figure 4−27. Principal component analysis of the bacteria α−diversity with soil physicochemical attributes and other environmental variables for WE−transect

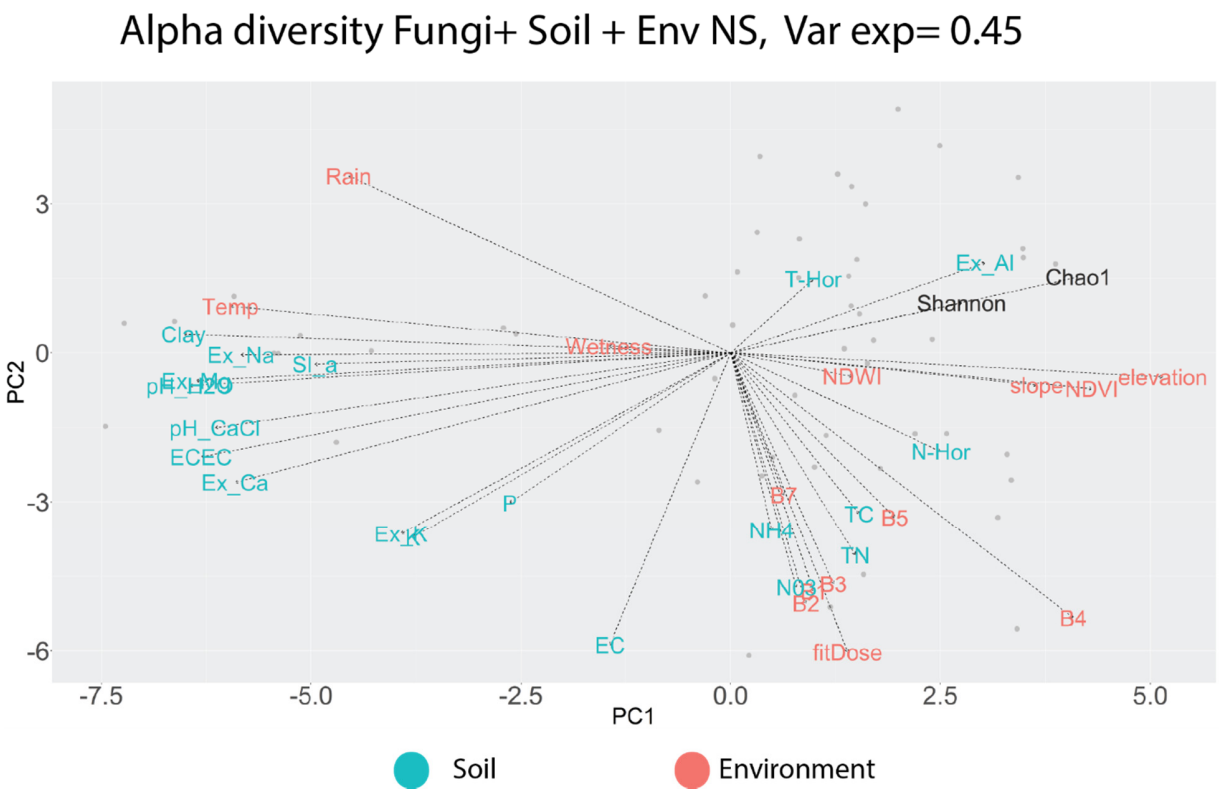


Figure 4-28. Principal component analysis of the fungi α-diversity with soil physicochemical attributes and other environmental variables for NS-transect

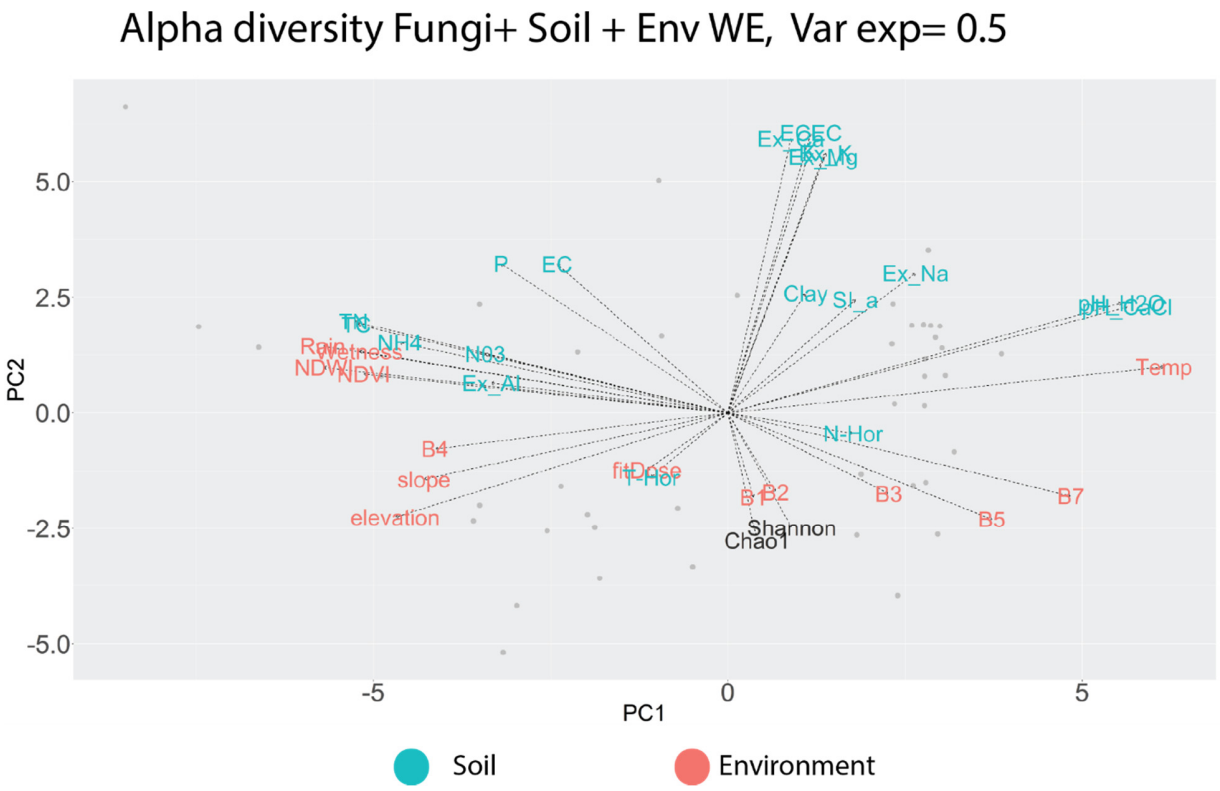


Figure 4−29. Principal component analysis of the fungi α−diversity with soil physicochemical attributes and other environmental variables for WE−transect

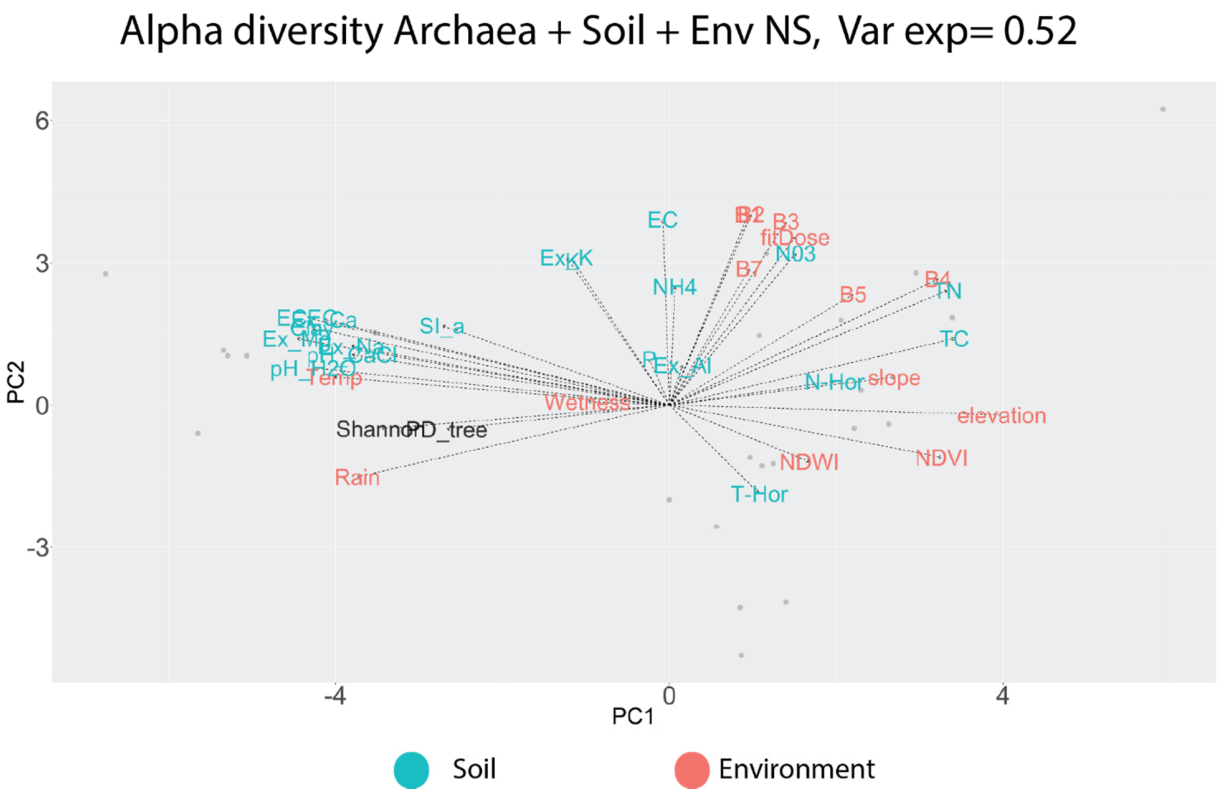


Figure 4-30. Principal component analysis of the archaea α-diversity with soil physicochemical attributes and other environmental variables for NS-transect.

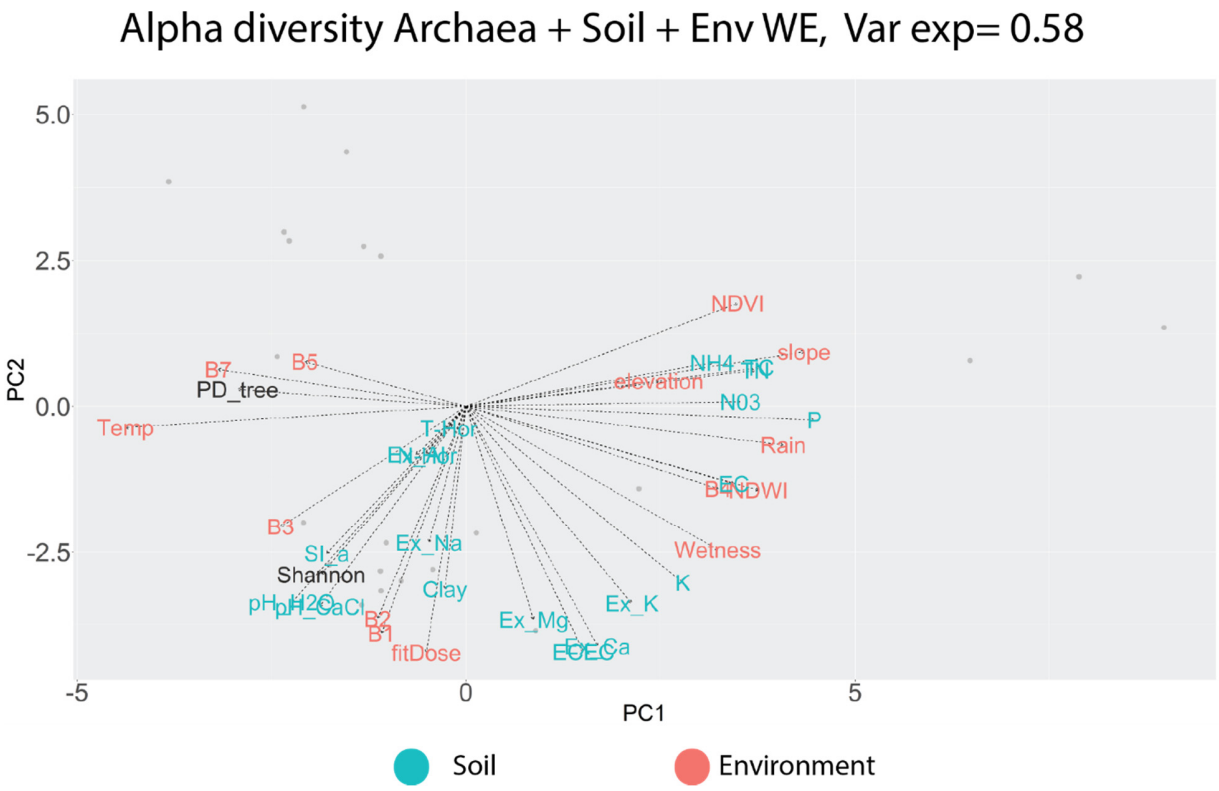


Figure 4−31. Principal component analysis of the archaea α−diversity with soil physicochemical attributes and other environmental variables for WE−transect.

MODELLING AND MAPPING SOIL MICROBIAL DIVERSITY

In order to analyse the geographical patterns of both pedological and environmental influence over soil microbial α-diversity, we performed a multivariate modelling of bacterial, fungal and archaeal diversity at 1 km pixel resolution in NSW. Microbial diversity of the three kingdoms was modelled using a 50 model bootstrap rule instance-based regression modelling approach (Kuhn et al., 2014; Quinlan, 1986). The selected models were validated using a totally external dataset (25% validation dataset). By means of the web platform of Google Earth Engine (Google Earth Engine Team, 2015) both laboratory measurements and remotely sensed data were organised as input variables for modelling. Such input variables were used as 25 (30 – 1,000 m pixel resolution) raster layers which included the most widely documented abiotic variables affecting microbial spatial distribution (Aschonitis et al., 2016; Dequiedt et al., 2009; Fierer et al., 2007; Fierer and Jackson, 2006; Green et al., 2004; Lauber et al., 2009; Maestre et al., 2015; Shahbazi et al., 2013; Xiong et al., 2012; Zhou et al., 2016). As a result, we obtained microbial diversity maps describing different patterns depending on the analysed microbial kingdom. Both modelling performance and predictions in this exercise were comparable with previous mapping efforts (Constancias et al., 2015a; Griffiths et al., 2015, 2011; Tedersoo et al., 2014).

The soil microbial diversity modelling was carried out by following the procedure below:

(i) Preparation of the microbial and environmental covariates dataset:

   a. Microbial diversity data (Chao1, Shannon) per each sampling site of the study

area.

b. Environmental covariates were selected in order of the following preference: > laboratory measurements> remote sensing data (rasters) and > seasonal variables (*e.g.* land surface temperature) according our field sampling campaigns seasons.

(ii) Split general dataset into *training* (75%) and *validation* (25%) datasets.

(iii) Generation models by bootstrap (modelling).

(iv) Selection models and microbial predictions.

(v) Soil microbial diversity mapping and model performance evaluation

Preparation of the microbial and environmental covariates dataset

To prepare the datasets including both soil microbial diversity data (diversity indices), soil attributes laboratory measurements and other remotely sensed environmental data we used the web platform Google Earth Engine (Google Earth Engine Team, 2015). All the information was uploaded to this platform for being organised and then exported as input variables for modelling analysis. Detailed protocols used to prepare each these datasets is described below.

Microbial input data from sampling sites

Regarding the microbial input data, each microbial models (i.e. bacteria, fungi and archaea) involved different number of valid samples in accordance with the microbes found across the sampling sites (e.g. archaea were found only in 33 of 49 sampling sites). This meant that the modelling datasets consisted of 94, 96 and 46 (out of a total of 98) valid samples for bacteria, fungi and archaea, respectively.

Environmental covariate selection

The environmental covariates used as input in our models were selected according to the following preferences. First, soil attributes that had both a valid laboratory observation and a corresponding raster coverage were extracted from the laboratory measurement. Such cases were clay (%), silt (%), sand (%), TN (%), pH and ECEC (cmol (+) kg^{-1}). Second, those soils attributes that had only raster coverage were extracted from the acquired rasters (ABARES, 2014; Farr and Kobrick, 2001; Grundy et al., 2015; Minty et al., 2009). These attributes were organic carbon (%), bulk density (Db) (gcm^{-3}), available water capacity (AWC) (%), total phosphorus (%), elevation (m), slope (degrees), salinity (presence/absence) and gamma radiometric (dose rate; Minty et al., 2009). Third, since the sampling campaigns for each of the transects were carried out in the same season but different years (see Chapter 2), those environmental attributes affected by seasonality (*e.g.* Land surface Temperature, Landsat bands and other their derivatives of them), were estimated as the median value of a set of observations in a pre-defined range dates according with the sampling campaigns. For example, Landsat 7 bands values were extracted as the median values of each raster cell between January and June 2012 for $NS_{\text{-transect}}$, and between

January and June 2013 for $WE_{-transect}$. In the end all the period dates were chosen based on both the sampling campaign dates and the availability of valid imagery (*e.g.* only images with less than 10% cloud coverage were included). For the special case of precipitation, we used a 10 year average due to inaccessibility to other sources (ABARES, 2014).

Once all the variables were extracted to the datasets, a final raster stack was compiled with all the previously mentioned variables plus Land surface Temperature and Landsat bands for the summer season 2016 (January to June 2016) in order to create a final map of predictions. All the data references and their sources are detailed in Table 5-7 (Appendix 4).

Split general dataset into *training* and *validation* datasets

The general datasets including microbial and environmental covariates was split into *training* and *validation* datasets using a Latin hypercube sample technique in R (Minasny and McBratney, 2006; Roudier et al., 2012). A 75% Latin hypercube sample was selected considering all of the input variables (25) plus the target soil microbial variables (*H' or Chao1*) as sources of variation. The Latin hypercube is a stratified random procedure which intends to maximize the distribution of the samples in a multivariate space. The technique depends directly on the number of random iterations of the algorithm and, for this reason, the procedure was performed with 15,000 iterations to ensure an adequate representation of the original dataset. The remainder 25% of the total valid samples was used as a totally independent validation dataset for microbial diversity modelling in downstream steps.

Generation models by bootstrap (modelling)

Even though Google Earth Engine is a platform that provides a suite of known modelling procedures (*e.g. inter alia* regression trees, artificial neural networks and support vector machines), bootstrapping and other resampling methods, commonly used for measuring predictions 'uncertainty', are still in an early stage of development (Padarian et al., 2015). Consequently, we used Google Earth Engine platform to stack together all the final rasters and once re-scaled they exported as a 25 layered raster stack at 1,000 m resolution for being used as input in the modelling process run in the High Performance Computing Facilities of The University of Sydney (The University of Sydney, 2016).

In order to have an indicator of the uncertainty of the predictions, the predicted values were the result of a 50-model bootstrap procedure on the training dataset for each of the predicted target variables, *i.e*. Chao1 and H' indices in each microbial kingdom modelled. Both the modelling and the predictions estimations were implemented in R (R Team et al., 2013).

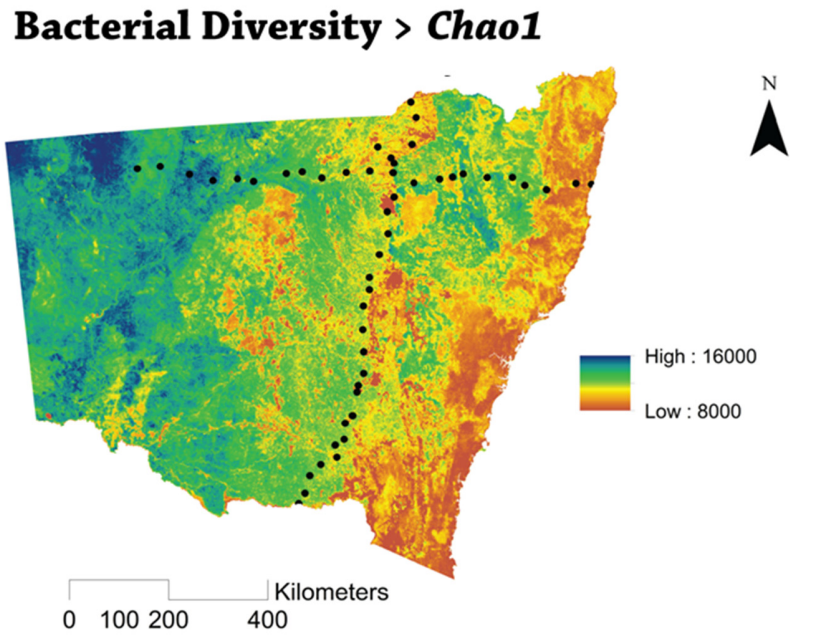
Selection of models and microbial predictions

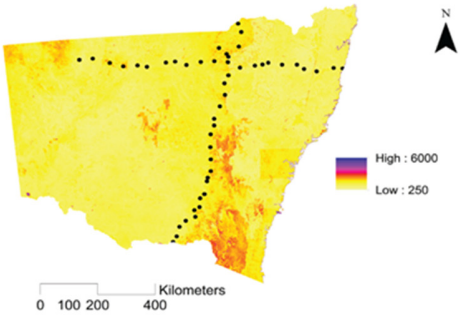
In order to assess the performance of each of the 50 models, 3 uncertainty coefficients were considered in this study *i.e.* R^2, RMSE and bias. After calculating the respective coefficients on each of the 50 models (*i.e.* by predicting against the validation dataset), the upper 2.5 percentile and the lower 25 percentile of the R^2 values were left aside, hence, only those models within the selected range were nominated to create the final predictions. This procedure was applied in order to both reduce the over fitting on the training data and avoid the use of less predictive models.

Once the models were selected, the respective regression splits (conditions that produced modelling branches) and regression tips (linear models at the end of a branch) were summed and recorded. The previous was presented as the total number of times that each of the variables in the model were used either as a regression split or as a regression tip. It is worth point out that most of the environmental variables were used in the tip regression models and therefore we decided to inform only the five most best ranked. For more details in regression trees and conditions or regression splits see (Kuhn et al., 2014).

Soil microbial diversity mapping and model performances evaluation

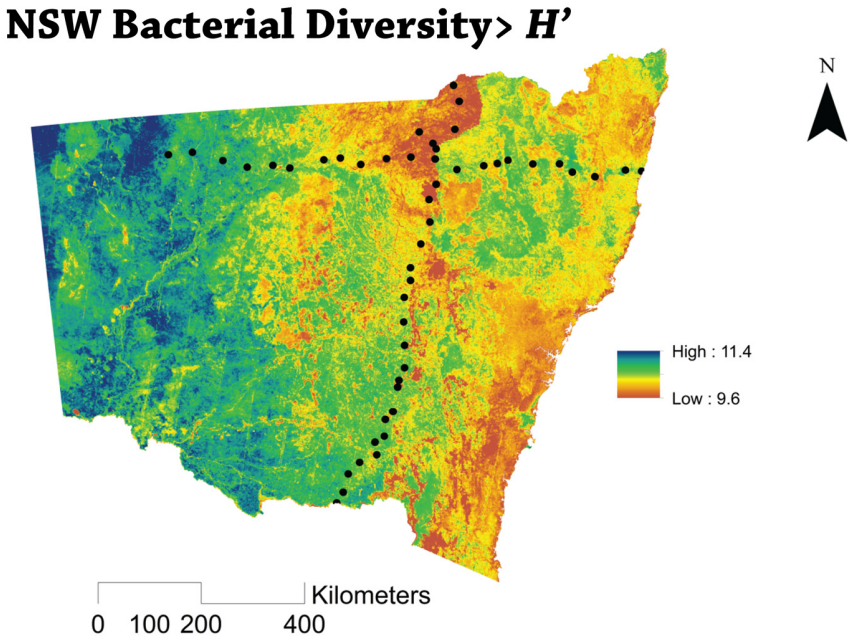
The models selected based on the previous method were applied over the 25 raster layered stack of NSW, and the mean results were compared against the observed samples to obtain the respective performance coefficients (R^2) for the training and validation datasets. Lastly, the prediction's standard deviation of the selected models calculated per pixel and mapped as the uncertainty of predictions. In this estimations the lowest standard deviations meant a robust modelling (less uncertainty). Finally, the resulting soil microbial diversity predicted distributions are shown mapped across NSW in terms of Chao1 and Shannon separated by each microbial kingdom in the following Figure 4-32 to Figure 4-37.

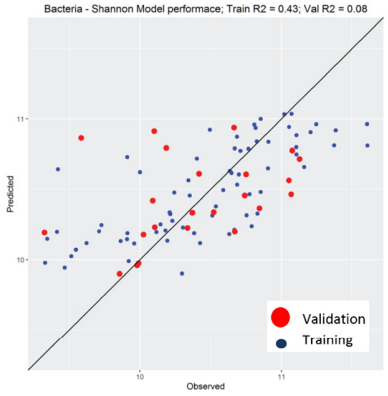


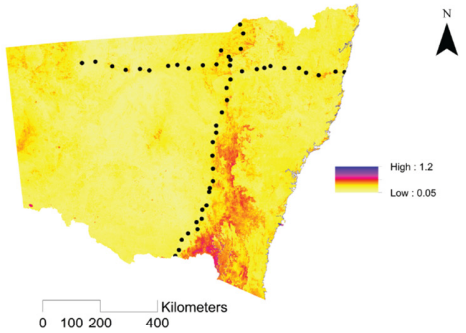


Variable[1]	Regression splits	Variable	Model tips
Precipitation	1630	B5	1635
B3	600	Sand	1328
Land Surface Temperature	156	Precipitation	1326
Organic carbon	107	Land Surface Temperature	1228
B1	100	Clay	1063

Figure 4-32. Mean predicted bacteria α-diversity (Chao1) distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.

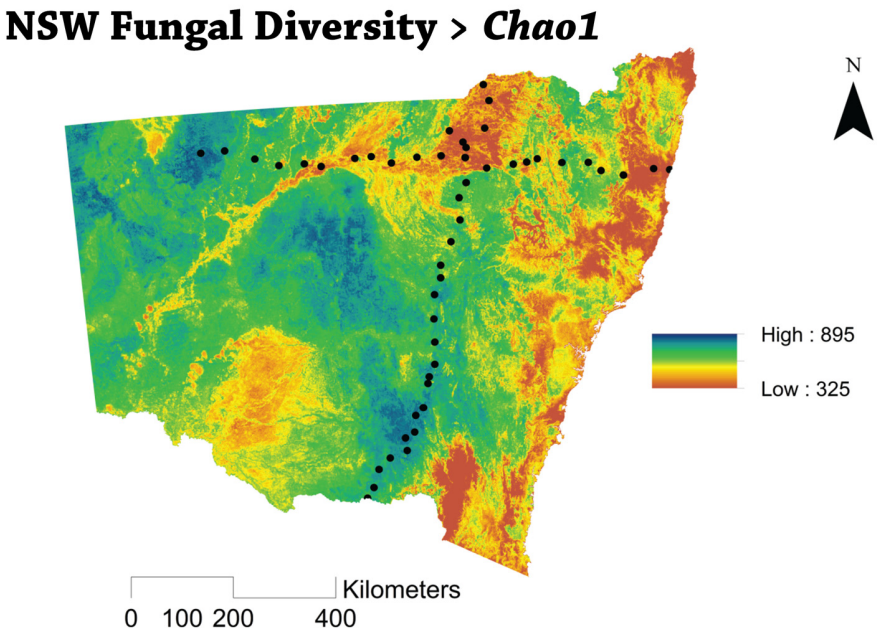


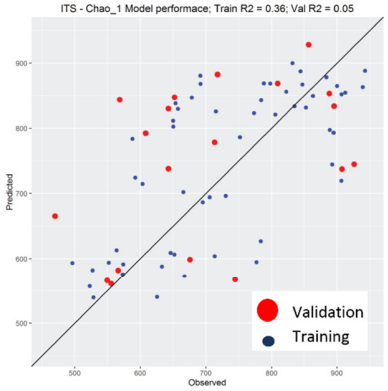


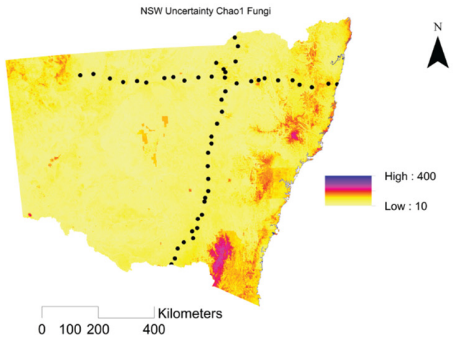


Variable	Regression splits	Variable	Model tips
Precipitation	1630	B5	1958
Silt	600	Organic carbon	1531
Bulk density	156	Precipitation	1463
Sand	107	Land Surface Temperature	1407
		B3	801

Figure 4-33. Mean predicted bacteria α-diversity (H') distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.

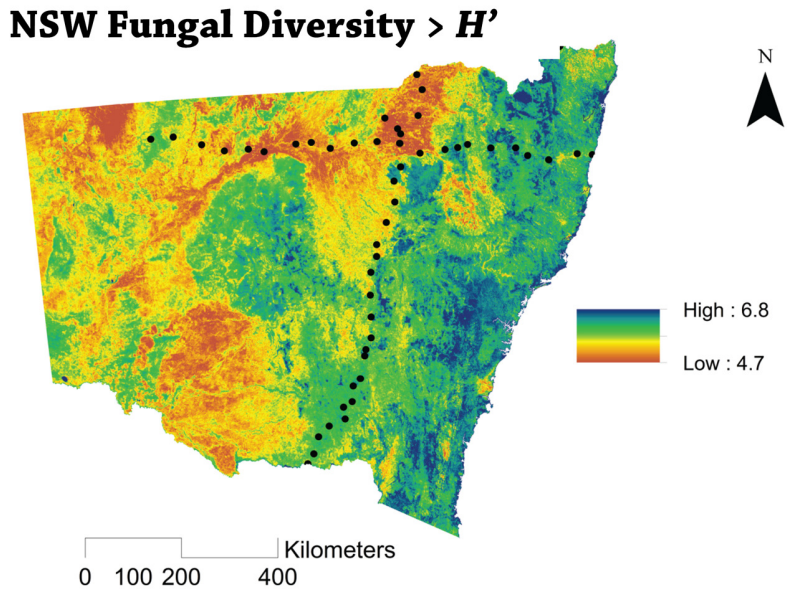


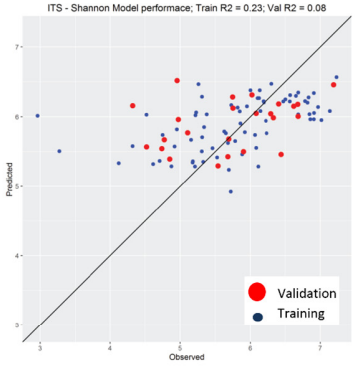


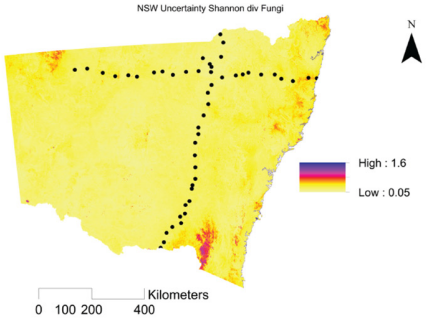


Variable	Regression splits	Variable	Model tips
ECEC	300	ECEC	2736
		Total nitrogen	2700
		Bulk Density	2400
		Wetness index	772
		pH	764

Figure 4-34. Mean predicted fungal α-diversity (Chao1) distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.

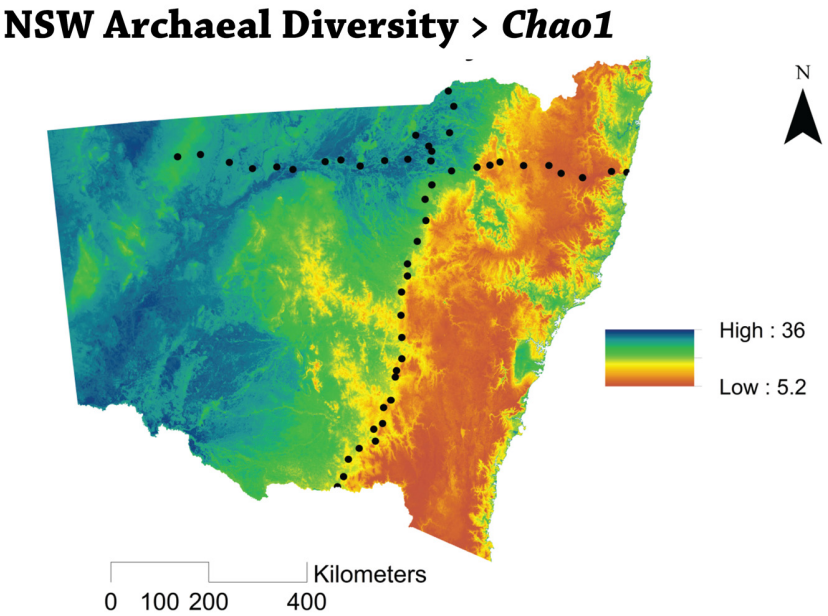


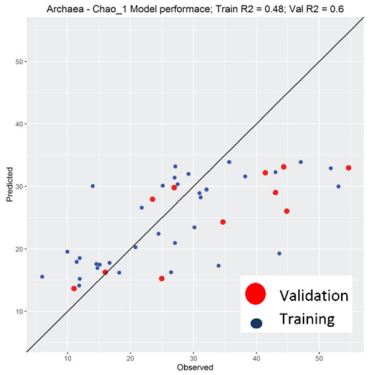


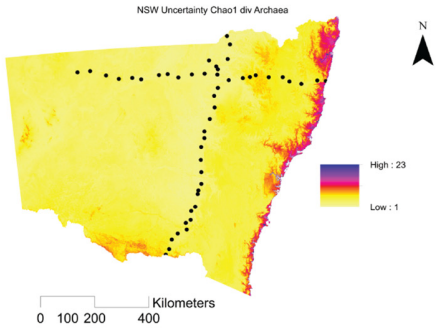


Variable	Regression splits	Variable	Model tips
ECEC	100	ECEC	2900
		B5	1000
		Land Surface Temperature	1000
		Organic carbon	800
		NDVI	600

Figure 4-35. Mean predicted fungal α-diversity (H') distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.

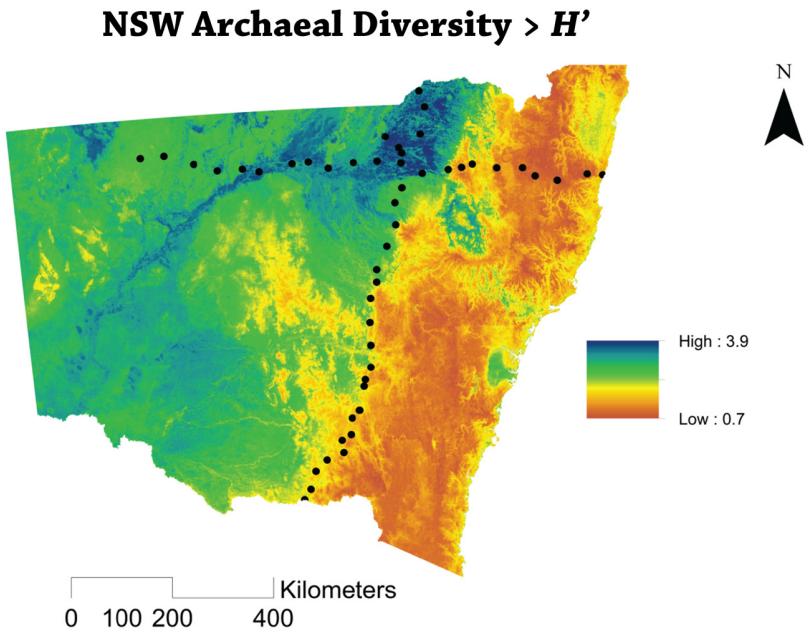


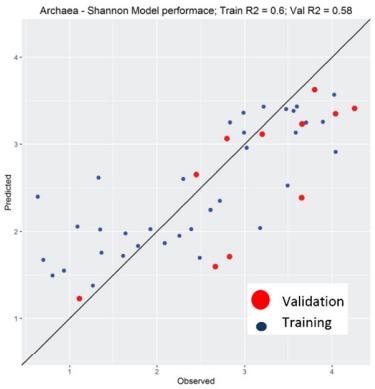


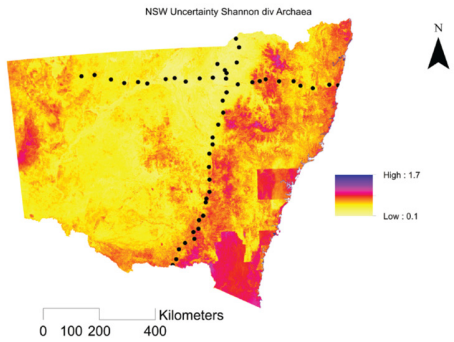


Variable	Regression splits	Variable	Model tips
Organic Carbon	100	Elevation	1400
		Organic carbon	1300
		pH	1000
		Precipitation	300
		ECEC	200

Figure 4−36. Mean predicted archaeal α−diversity (Chao1) distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.







Variable	Regression splits	Variable	Model tips
Organic Carbon	100	Elevation	1400
		Organic carbon	1300
		pH	1000
		Precipitation	300
		ECEC	200

Figure 4-37. Mean predicted archaeal α-diversity (H') distribution across NSW. Model performance (R^2 of predicted vs observed microbial value) and uncertainty map (standard deviation of predictions) are shown.

Spatial diversity pattern of the predicted soil microbial diversity in NSW

Six soil microbial diversity predictions maps were obtained by modelling Chao1 and H' diversity of each microbial kingdom. In general terms, as expected, the main attributes shaping microbial distribution were variable per microbial kingdom, although, some general patterns were also clear. Based on this results we state that soil microbial diversity in NSW revealed geographical patterns. The most notorious pattern in the first place is the effect of the physiographic features of NSW (see Chapter 2, Figure 2-4). In the context of these physiographic divisions, there is an evident very low level of soil microbial diversity in the *Eastern Upland Division* zones, eastern of NSW. Indeed, as it was also indicated in Chapter 2, this physiographic features signify one of the most important differentiations between the distributions of NSW soil types; regolith material limited to *Western zones* (*Interior Lowlands* and *Wester Plateau*) whereas *Eastern Uplands* areas mostly comprised of sedimentary materials (ABARES, 2014; Blewett, 2012).

In general, we obtained very consistent results between linear correlations, PCAs and modelling regressions. These consistencies can be seen more clearly by analysing the maps together with the description of the models performed during the mapping. For example, there were seen in PCA and correlation analysis that *archaea* and *fungi* resulted more closely related with soil attributes than *bacteria*. This same pattern was observed by modelling, in which, both *archaea* and *fungi* were primary split by Organic Carbon (100 regression splits for Chao1; Figure 4-37) and ECEC (300 regression splits for Chao1; Figure 4-34), respectively. But bacterial diversity had

a more complex performance and resulted being primary split by many different soil and environmental variables in which precipitation was the best ranked (1630 regression splits: Figure 4-32 and Figure 4-33).

A major description of the performance carried out during the modelling process by each microbial group is discussed below.

Bacterial diversity predicted across NSW

Bacterial diversity (H' and Chao1) modelling used a training dataset of 47 samples based on which were created 50 models. 35 out of 50 models were used to create the final predictions. Figure 4-32 and Figure 4-33 illustrates the number of times each input variable was used as a main split on those 35 models for Chao1 and H' microbial diversity predictions.

It is important to remember that the observed relations are the result of non-linear relations. Hence, and as described before, the complex relationship between single either soil or environmental attributes are not necessarily the exactly the same observed in the linear correlations and/or principal component analyses. From the tables observed in Figure 4-32 and Figure 4-33, we highlight the high complexity in the distribution patterns of bacterial diversity. A first glance shows a negative longitudinal gradient from west to east followed both precipitation and temperature gradients. However, there is a mix effect seen in terms of longitudinal gradients. For example, we used to describe lower bacterial diversity toward the northern part of $NS_{\text{-transects}}$ when encompassing *Vertosols* and *Kandosols*, represented by the set of

attributes we more discussed earlier analysing PCAs, *i.e.* clay, Ex_Na, SI, ECEC, etc. However, our maps shows that this pattern change completely if we move this transect few km to the east. In this sense, it is difficult to state, for examples, if bacterial diversity increases towards the tropics or not. What is more appropriate to say is that depending on the longitudinal position the 'Clay/Sand (model tips)' represented higher/lower diversity in these maps (especially in *Vertosols*). The most clear bacterial pattern was effect of temperature and precipitation both in the regression splits variables.

The uncertainty related with the models is presented as the standard deviation of the selected models predictions, in the case of *bacteria* diversity, the standard deviation of 35 different predictions (35 models used for Shannon and Chao 1). It can be seen that the predictions were sufficiently robust in most of the area; however in those regions with higher environmental complexity (*e.g.* landforms) the uncertainty was higher. It is important to highlight that, even we believe the high importance of reporting the uncertainty of a prediction, this type of analysis is not usually used for microbial diversity studies. We believe that these modelling procedure increases the value of this investigation.

Soil fungal diversity patterns across NSW

By following the physiographic divisions in this map we can see how the fungal distribution in our diversity maps follow a quite similar pattern. As described in Chapter 2, these physiographic provinces are divided according to their similarities of landforms features (slope, relief) but also in term of soil orders and water balance. As we discussed below, also fungi was the microbial group primary predicted by soil attributes Figure 4-34 to Figure 4-35.

In the fungi diversity modelling, the training dataset comprised 72 and 24 training and validation samples respectively. The results showed a clear and consistent influence of the soil type. This pattern was immediately associated with the distribution of soil order ASC across NSW (see Chapter 2, ASC). The relations between fungal diversity and soil types were evident with *Vertosols*, *Kandosols*, *Calcarosols* and *Sodosols*. Accordingly, the lowest diversity were associated with both lower clay content and lower soil pH. Both attributes were also informed in correlation analysis (Figure 4-23**Error! Reference source not found.**). Similarly, the regression splits in our models were dominated by ECEC and other 5 highest ranked linear regression tips, *i.e.* ECEC, TN, bulk density, WI and soil pH (Figure 4-34 to Figure 4-35).

This study showed the lower diversity in soils with heavier textures and low aggregation features, which are coincidently with *Vertosols* in norther NSW. These soils were also the ones with higher SI, *i.e.* the most instable of soils. These results seems to agree with the statements of (Peay et al., 2016; Six et al., 2004; Vos et al., 2013) relating higher aggregation or stability to hyphal interaction between soil fungal communities and the soil substrate.

The dispersion of the predicted values (Figure 4-34 to Figure 4-35), once again showed a low predictability, especially in the validation dataset (R^2 0.05) which evidence that other factors not considered in this study might be involved in the fungal diversity. Despite of the previous, it is important to point out that the results of both linear (correlations and principal component analysis) and non- linear analyses led to similar results showing the same relations between pH, and Cation Exchangeable Capacity.

Soil archaeal diversity patterns across NSW

Archaea's training dataset comprised 35 and 11 training and validation samples respectively. It was observed, as expected from the lower amount of valid samples (hence, spatial coverage) that the geographical patterns capture by the decision trees modelling was also very general. Again a simple condition rule of content of Organic carbon (%) was created by all the 33 models selected (out of 50).

The linear regression models considered the same property plus elevation, which is also noticeable in the prediction map, where a clear cut between the western plains and the mountain range showed a strong west-east gradient (Figure 4-36 and Figure 4-37). In terms of soil types, there was a notorious contrasting relation between a*rchaea* and *fungi*. In this sense, we found higher diversity of *archaea* (both Shannon and Chao 1) in clayey soils (*Vertosols*) rather than coarser texture soils (*Kandosols*). This results were quite consistent with PCAs analysis (Figure 4-24).

Despite of the lowest diversity and abundances ranges of *archaea*, out of the three kingdom's diversities modelled, they had the best performance. These results may suggest that either the factors controlling *archaea* diversity less complex than those controlling the other kingdoms or that these results are a reflection of the fewer taxonomic classes present in this taxa.

REFERENCES

ABARES, 2014. 2014 Datapack to support the Multi-Criteria Analysis Shell for Spatial Decision Support (MCAS-S).

Aschonitis, V.G., Castaldelli, G., Fano, E.A., 2016. Relations between environmental gradients and diversity indices of benthic invertebrates in lotic systems of northern Italy. Web Ecol. 16, 13–15. doi:10.5194/we-16-13-2016

Bâ, a. M., Diédhiou, a. G., Prin, Y., Galiana, a., Duponnois, R., 2010. Management of ectomycorrhizal symbionts associated to useful exotic tree species to improve reforestation performances in tropical Africa. Ann. For. Sci. 67, 301–301. doi:10.1051/forest/2009108

Blewett, R., 2012. Shaping a nation : a geology of Australia, 34th International Geological Congress.

Cao, P., Zhang, L.M., Shen, J.P., Zheng, Y.M., Di, H.J., He, J.Z., 2012. Distribution and diversity of archaeal communities in selected Chinese soils. FEMS Microbiol. Ecol. 80, 146–158. doi:10.1111/j.1574-6941.2011.01280.x

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

Nat. Publ. Gr. 7, 335–336. doi:10.1038/nmeth0510-335

Castañeda, L.E., Godoy, K., Manzano, M., Marquet, P.A., Barbosa, O., 2015. Comparison of soil microbial communities inhabiting vineyards and native sclerophyllous forests in central Chile. Ecol. Evol. 5, 3857–3868. doi:10.1002/ece3.1652

Constancias, F., Saby, N.P.A., Terrat, S., Dequiedt, S., Horrigue, W., Nowak, V., Guillemin, J.P., Biju-Duval, L., Chemidlin Pr??vost-Bour??, N., Ranjard, L., 2015. Contrasting spatial patterns and ecological attributes of soil bacterial and archaeal taxa across a landscape. Microbiologyopen 518–531. doi:10.1002/mbo3.256

Dequiedt, S., Thioulouse, J., Jolivet, C., Saby, N.P. a, Lelievre, M., Maron, P.A., Martin, M.P., Prévost-Bouré, N.C., Toutain, B., Arrouays, D., Lemanceau, P., Ranjard, L., 2009. Biogeographical patterns of soil bacterial communities. Environ. Microbiol. Rep. 1, 251–255. doi:10.1111/j.1758-2229.2009.00040.x

Farr, T., Kobrick, M., 2001. The shuttle radar topography mission. Eos Trans. AGU 82, 47. doi:10.1029/2005RG000183.1.INTRODUCTION

Fenchel, T., Finlay, B.J., 2004. The Ubiquity of Small Species: Patterns of Local and Global Diversity. Bioscience 54, 777. doi:10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proc. Natl. Acad. Sci. U. S. A. 103, 626–631. doi:10.1073/pnas.0507535103

Fierer, N., Morse, J.L., Berthrong, S.T., Bernhardt, E.S., Jackson, R.B., 2007. Environmental controls on the landscape-scale biogeography of stream bacterial communities. Ecology 88, 2162–2173. doi:10.1890/06-1746.1

Fuhrman, J. a, 2009. Microbial community structure and its functional implications. Nature

459, 193–199. doi:10.1038/nature08058

Google Earth Engine Team, 2015. Google Earth Engine: A planetary-scale geospatial analysis platform.

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M., Beattie, A.J., 2004. Spatial scaling of microbial eukaryote diversity. Nature 432, 747–750. doi:10.1038/nature03034

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils. Environ. Microbiol. 13, 1642–1654. doi:10.1111/j.1462-2920.2011.02480.x

Griffiths, R.I., Thomson, B.C., Plassart, P., Gweon, H.S., Stone, D., Creamer, R.E., Lemanceau, P., Bailey, M.J., 2015. Mapping and validating predictions of bacterial biodiversity using European and national scale datasets. Appl. Soil Ecol. doi:10.1016/j.apsoil.2015.06.018

Grundy, M.J., Viscarra Rossel, R.A., Searle, R.D., Wilson, P.L., Chen, C., Gregory, L.J., 2015. Soil and landscape grid of Australia. Soil Res. 53, 835–844. doi:10.1071/SR15191

Gu, J., Mitchell, R., 2006. The Prokaryotes. doi:10.1007/0-387-30741-9

Hartmann, M., Lee, S., Hallam, S.J., Mohn, W.W., 2009. Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. Environ. Microbiol. 11, 3045–3062. doi:10.1111/j.1462-2920.2009.02008.x

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Kotaro, I., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree and life's diversity. Nat. Microbiol. 1, 16048. doi:10.1038/nmicrobiol.2016.48

Kasel, S., Bennett, L.T., Tibbits, J., 2008. Land use influences soil fungal community composition across central Victoria, south-eastern Australia. Soil Biol. Biochem. 40, 1724–1732. doi:10.1016/j.soilbio.2008.02.011

Kuhn, M., Weston, S., Keefer, C., Coulter, N., 2014. Cubist: Rule and Instance-Based Regression Modeling. R package version 0.0.18.

Kuramae, E.E., Hillekens, R.H.E., de Hollander, M., van der Heijden, M.G.A., van den Berg, M., van Straalen, N.M., Kowalchuk, G.A., 2013. Structural and functional variation in soil fungal communities associated with litter bags containing maize leaf. FEMS Microbiol. Ecol. 84, 519–531. doi:10.1111/1574-6941.12080

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75, 5111–5120. doi:10.1128/AEM.00335-09

Lauber, C.L., Strickland, M.S., Bradford, M. a., Fierer, N., 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol. Biochem. 40, 2407–2415. doi:10.1016/j.soilbio.2008.05.021

Ma, A., Zhuang, X., Wu, J., Cui, M., Lv, D., Liu, C., Zhuang, G., 2013. Ascomycota Members Dominate Fungal Communities during Straw Residue Decomposition in Arable Soil. PLoS One 8, 1–9. doi:10.1371/journal.pone.0066146

Maestre, F.T., Delgado-Baquerizo, M., Jeffries, T.C., Eldridge, D.J., Ochoa, V., Gozalo, B., Quero, J.L., García-Gómez, M., Gallardo, A., Ulrich, W., Bowker, M.A., Arredondo, T., Barraza-Zepeda, C., Bran, D., Florentino, A., Gaitán, J., Gutiérrez, J.R., Huber-Sannwald, E., Jankju, M., Mau, R.L., Miriti, M., Naseri, K., Ospina, A., Stavi, I., Wang, D., Woods, N.N., Yuan, X., Zaady, E., Singh, B.K., 2015. Increasing aridity reduces soil microbial diversity and

abundance in global drylands. Proc. Natl. Acad. Sci. U. S. A. 112, 15684–15689. doi:10.1073/pnas.1516684112

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Publishing Ltd, Oxford, UK.

McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Minty, B., Franklin, R., Milligan, P., Richardson, M., Wilford, J., 2009. The Radiometric Map of Australia. Explor. Geophys. 40, 325–333. doi:10.1071/EG09025

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T.S., Meiners, T., M??ller, C., Obermaier, E., Prati, D., Socher, S.A., Sonnemann, I., W??schke, N., Wubet, T., Wurst, S., Rillig, M.C., 2014. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. Ecol. Evol. 4, 3514–3524. doi:10.1002/ece3.1155

Offre, P., Spang, A., Schleper, C., 2013. Archaea in biogeochemical cycles. Annu. Rev. Microbiol. 67, 437–57. doi:10.1146/annurev-micro-092412-155614

Padarian, J., Minasny, B., McBratney, A.B., 2015. Using Google's cloud-based platform for digital soil mapping. Comput. Geosci. 83, 80–88. doi:10.1016/j.cageo.2015.06.023

Peay, K.G., Kennedy, P.G., Talbot, J.M., 2016. Dimensions of biodiversity in the Earth mycobiome. Nat. Rev. Microbiol. 14, 434–447. doi:10.1038/nrmicro.2016.59

Pielou, E.C., 1975. Ecological diversity. John Wiley Sons VIII, 165p.

Quinlan, J.R., 1986. Induction of Decision Trees. Mach. Learn. 1, 81–106.

doi:10.1023/A:1022643204877

Rampelotto, P.H., de Siqueira Ferreira, A., Barboza, A.D.M., Roesch, L.F.W., 2013. Changes in Diversity, Abundance, and Structure of Soil Bacterial Communities in Brazilian Savanna Under Different Land Use Systems. Microb. Ecol. 66, 593–607. doi:10.1007/s00248-013-0235-y

R Team, C., others, R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Rime, T., Hartmann, M., Brunner, I., Widmer, F., Zeyer, J., Frey, B., 2015. Vertical distribution of the soil microbiota along a successional gradient in a glacier forefield. Mol. Ecol. n/a-n/a. doi:10.1111/mec.13051

Roudier, P., Hewitt, A.E., Beaudette, D.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. Digit. Soil Assessments Beyond 227–231. doi:10.1201/b12728-46

Schleper, C., Nicol, G.W., 2010. Ammonia-oxidising archaea - physiology, ecology and evolution. Adv. Microb. Physiol. 57, 1–41. doi:10.1016/B978-0-12-381045-8.00001-1

Shahbazi, F., Aliasgharzad, N., Ebrahimzad, S. a., Najafi, N., 2013. Geostatistical analysis for predicting soil biological maps under different scenarios of land use. Eur. J. Soil Biol. 55, 20–27. doi:10.1016/j.ejsobi.2012.10.009

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31. doi:10.1016/j.still.2004.03.008

Smith, B., Wilson, J.B., A, J.B., 1996. A Consumer ' s Guide to evenness indices 76, 70–82.

Spain, A.M., Krumholz, L.R., Elshahed, M.S., 2009. Abundance, composition, diversity and

novelty of soil Proteobacteria. ISME J. 3, 992–1000. doi:10.1038/ismej.2009.43

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, S., Wardle, D.A., Lindahl, B.D., 2014. Global diversity and geography of soil fungi. Science (80-.). 346, 1052–1053. doi:10.1126/science.aaa1185

The University of SYDNEY, 2016. Artemis: The University of Sydney High Performance Facilities.

Uroz, S., Ioannidis, P., Lengelle, J., Cébron, A., Morin, E., Buée, M., Martin, F., 2013. Functional Assays and Metagenomic Analyses Reveals Differences between the Microbial Communities Inhabiting the Soil Horizons of a Norway Spruce Plantation. PLoS One 8, e55929. doi:10.1371/journal.pone.0055929

Valverde, A., Tuffin, M., Cowan, D. a., 2012. Biogeography of bacterial communities in hot springs: A focus on the actinobacteria. Extremophiles 16, 669–679. doi:10.1007/s00792-012-0465-9

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol. Mol. Biol. Rev. 71, 495–548. doi:10.1128/MMBR.00005-07

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G. a., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev. 37, 936–954. doi:10.1111/1574-6976.12023

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., Yang, Y., Yao, T., Knight, R., Chu, H., 2012. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. Environ. Microbiol. 14, 2457–2466. doi:10.1111/j.1462-2920.2012.02799.x

Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., Qin, Y., Xue, K., Wu, L., He, Z., Voordeckers, J.W., Nostrand, J.D. Van, Buzzard, V., Michaletz, S.T., Enquist, B.J., Weiser, M.D., Kaspari, M., Waide, R., Yang, Y., Brown, J.H., 2016. Temperature mediates continental-scale diversity of microbes in forest soils. Nat. Commun. 7, 12083. doi:10.1038/ncomms12083

Chapter 5

Conclusions & Future Work

Outlines

This is our final discussion and conclusion chapter. We present a final discussion based on the main findings found all along this investigation including those about methodological decisions. The discussion is conducted by revisiting each chapter in which some specific conclusions are also provided. In order to summarize important results on which we based the final conclusions, we provide a deeper analysis into Chapter 4. Finally, there are some directions about pending and further works for the final achievement of the main hypothesis presented beyond this investigation.

FINAL DISCUSSION

Chapter 1

In the process of writing this thesis, we discovered a surprising dearth of information relating soil microbial patterns and other soil attributes from a pedological perspective. We also found that the majority of scientific investigations concerning this biotic-abiotic interaction at larger spatial extents were skewed towards microbiological and ecological explorations. More exhaustive studies from a pedological perspective have usually been carried out in the form of edaphological position such as by assessing the influence of soil properties on living organisms at local scales. We drew significantly upon the information contained in these other studies in order to respond to the first question which we posed for our hypothesis; that the distribution of soil microbes around the globe are controlled by soil attributes.

Our proposal relies on the most basic concept of soil formation processes described by Jenny in 1941, i.e. almost a hundred years ago. Jenny outlined the essential soil-forming factors shaping our soil environment, i.e. *climate*, *organisms*, *topography*, *parent material* and *time*. In soil science, we know the extent of soil physiochemical properties correlated with soil-forming factors, and this knowledge became a key approach when analysing the differentiation of soil composition and their distribution. Later, soils across the world began to be mapped in order to assist the understanding of their spatio-temporal differentiations and, since then, two new soil-forming factors were proposed by McBratney et al., (2003): *space/spatial position* and *soil* - the *soil* itself has

been included as a factor for the modelling of its own spatial distribution since it can be predicted by its properties (McBratney et al., 2003). As revised in Chapter 1, the soil-forming factors above encompasses all the principal elements found in most of the current investigations searching for main drivers shaping soil microbial communities patterns within any given spatiotemporal scale. On this fact we rely our notion of the co-spatial relations between soil pedodiversity (and/or entities i.e. classes, horizons) and biodiversity. In this thesis, we began testing this relationship firstly confirming that the soil properties in our study area are intimately related with the patterns of the soil microbial communities.

Certainly, as far as we understand soil ecosystems, i.e. one of the most complex and heterogeneous of all the terrestrial ecosystems, the mapping and prediction of their spatiotemporal distributions has not been easy for soil scientists. However, during the process of dealing these difficulties we have developed principles which help us to take decisions. For example, for years the process of soil mapping had deal with the fact that multivariable factors are controlling 'the *heterogeneity* of another *heterogeneous* factor'. This experience and *know-how* can be useful when analysing and deducting the complex relation of soil microbes in the context of their soil environments.

From this understanding more recently upsurged the notion of *pedodiversity-biodiversity* relationship in which is assumed the close relation between the diversity of both soil biota and other soil physicochemical attributes. From our point of view, this means that none of them are dependent on only one single attributes of each other. Certainly, there is no microbial diversity/abundance correlated with only one single soil properties but with a group of them -

which might be representing a degree of soil composition -. Commonly, these kind of exploratory analyses tend to find only one strong/significant relation (e.g. microbial abundance being driven by the content of total carbon) on which focus their analysis. However, from our perspective, this kind of analysis could be masking other important patterns and valuable information when interpreting microbial patterns.

Moreover, the degree of 'soil diversity' or 'pedodiversity', which is in fact defined by a set of soil properties, will determine the soil *spatiotemporal* differentiation that is used to categorised soil entities, e.g. soil horizons, classes. Consequently, the *pedodiversity-biodiversity* relationship results in a multidimensional functional unit. Upon this multidimensional-component is that rely the status of critical biogeochemical reactions in which are determined the soil functions ensuring soil ecosystem services (e.g. food production). Ecosystems services that are in fact widely recognized and cited in the literature (soil food web properties, nutrient cycling, C sequestration, water/air filtering, bioprospecting applications, genetic seed bank, etc.). This *pedodiversity-biodiversity* relationship is a multifunctional task and takes place in a complex environment. Today, there is an opportunity to gain a better understanding since the disciplines closely involved in developing this knowledge are becoming closer (e.g. microbiology, soil science).

This means that we will be able to understand the multidimensional variability of this pedodiversity-*biodiversity* dynamisms and, by doing so, we would provide 'metrics' by which to valuate soil ecosystems services and, therefore, to establish regulatory policies for their protection, e.g. *if we enable metrics for the soil biodiversity quantification we would enable its*

protection. These days, the study of *pedodiversity-biodiversity* interaction is emerging as the newest concept of *biopedometrics*.

The quantification/qualification of soil living organisms and the estimation of their contribution on the soil physicochemical processes is difficult but new scientific approaches are providing highly defined information. In this sense, the evaluation of soil physicochemical characteristics is less demanding especially with the use of newer instrumentations/approaches (e.g. Vis-NIR, remote sensing). Anyhow, there are different disciplines and initiatives providing valuable information beyond this soil biodiversity-pedodiversity compound – e.g. the Global Soil Biodiversity Initiative and the Global Digital Soil Mapping – but they are still working apart from each other and so yet distant from what the extent of a *biopedometrics* discipline would demand. Anyways, and as never before, the scientific community is attentive to gain a better understanding in this regard. For instance, it was recently launched a publication called '*Back to the future of soil metagenomics'* whose authors, a group of forty-five scientists from different areas (geneticist, soil scientists, ecological microbiology, etc.), expressed their interest in developing uniform methodological frameworks for the appropriate investigation of soil microbial communities, e.g. uniform soil DNA extraction protocols (Nesme et al., 2016). Furthermore, these authors manifested as common interest the necessity of generating a comprehensive catalog of all microbial community members and functions for at least one reference soil. This would generate valuable information that together with Soil Digital Mapping *know-how* could shed light on our still lack of understanding of soil microbial community patterns. In Australia, there has been recently introduced *The Biome of Australian Soil Environments (BASE),* a platform that currently

provides amplicon sequences and associated contextual data for over 900 sites encompassing all Australian states and territories. Perhaps a still missing call is the one we are trying to expose by means of this investigation, i.e. let's try to characterise soil microbial communities in the context of their own soil 'entities' or environments, i.e. *we already know the physicochemical differences of these 'soil habitats' (soil types, classes, horizons, gradients) and so let's try now to differentiate their inhabitants*.

This information can be useful for research disciplines as such as for other commercial areas, e.g. bioprospecting applications. Indeed, nowadays all the knowledge generated by *pedometrics* has provided important scientific support in *precision agriculture* – e.g. the adaptation of digital soil mapping for precision agriculture (Söderström et al., 2016). In the same way, the understanding and measuring of microbial communities' patterns would provide tool and informative data to be uses in the optimization of cropping managements.

For a long time, soil scientists have been strangers to the extent of the contribution of soil microorganisms for soil ecosystems as much as soil microbiologists have been to the extent of pedometrics approaches understanding the soil matrix. In this study we aimed to fill this gap by analysing soil microbial structural patterns from another perspective, one we believe is the first key factor for estimating the co-spatial pattern between these two biotic-abiotic components. Throughout this study we also aimed to provide some methodological advices such as the use of *conditioned Latin Hypercube method for the modelling* of soil microbial communities.

We concluded from Chapter 1 that there is a need to spatially correlate both soil physiochemical attributes and microbial geographical pattern distributions as much as to evaluate the potential beyond the use of biomolecular/HTS technology in conjunction with newest global soil digital mapping approaches.

Chapter 2

In Chapter 2, we presented our experimental design. In this regards we can recognise the convenience of analysing two orthogonal latitudinal and longitudinal gradients. In the first place, NSW (comparable in size to France) presented distinctive features in both longitudinal and latitudinal environmental gradients. These features were a useful guideline for data interpretation (e.g. cracking soils, humid and forested eastern areas). For instance, their orthogonality made possible to mitigate the effect of certain environmental attributes, e.g. precipitation. In this sense, since $NS_{\text{-transect}}$ kept the rainfall pattern constant this variable showed orthogonality to our microbial diversity variables. In the $WE_{\text{-transect}}$, since temperature and rainfall followed contrasted gradients across NSW, we were able to observe that cracking soil conditions were the main determinants of microbial patterns.

The use of two orthogonal transects also pointed out the fact that if we would decide to sample one or another, instead of those two, we would infer totally different results. We uncovered from the modelling analysis that if we would place $NS_{\text{-transect}}$ a few km in an easterly direction, our tendencies of increasing *fungi* and *bacteria* diversity towards the south would be opposite to these

findings. So there are evident risks in drawing conclusions from studies along single supposed environmental gradients.

The transects were comparable to other studies and, in fact, we were able to cover from local to landscapes, and as far as a regional scales. For modelling and predictions, this experimental design provided valuable information by covering the most distinctive environmental aspects of NSW. Additionally, $WE_{\text{-transect}}$ proximity to sealed roads gave us the opportunity to increase our ranges of environmental gradients allowing the sampling in temperate and dry areas of western of NSW where microbial diversity tended to be higher.

Another valuable contribution in this investigation has been the compilations of a considerable amount of digital high quality environmental data. Moreover, we had access to other additional sources of information in which in the near future we would be able to deepen our investigation on microbial communities' patterns. For example, we are interested to asses these patterns in regard to certain agricultural managements and/or other practices, e.g. we would be able to analyse the microbial association with specific coniferous areas using Landsat bands. There is still valuable data that will be useful in further analysis (e.g. β-diversity in a future work).

Chapter 3

In Chapter 3, we described the materials and methods to obtain both the soil microbial identification and the soil physicochemical dataset. Microbial identification based on 16S and ITS

rDNA sequencing characterisation resulted in a robust method for microbial identification. We obtained a dataset comprising 11,557,499 sequences distributed through 423,740 OTUs. Our decision for a paired-end DNA metabarcoding approach resulted gave us a robust and good quality dataset for our analyses. The quality of these data reduced our doubt and questioning regarding bias contamination. This was particularly important when analysing the rare community which had a determinant participation for some of our results and conclusions. Despite the low quality obtained in long sequences reads, we were able to access a robust dataset of sequences ~200bp.

In regards to the construction of the soil physicochemical dataset, we included three recently introduced estimators of soil physical attributes: thickness of the first horizon (*T-Hor*), spectrally derived number of horizons (*N-Hor*) and slaking index (SI). These estimators were surprisingly well correlated with most of the microbial estimators. These results suggested that the use of Vis-NIR instrumentation contributed to complete our set of laboratory measurements with these novel indicators. In total we measured 19 soil physicochemical properties. At the same time, this investigation was useful by testing them with a robust number and wide range of soil types.

The soil cores extracted from 100 cm depth obtained from the study area will be processed in further analysis together with the soil samples extracted from 5-10cm since the purpose of this thesis was only focused on the top soil layer, i.e. 0-5 cm depth.

From chapter three we conclude that the design of our DNA sequencing library resulted in good quality data to accomplish the aims of our investigation. Additionally, this library is a valuable

source of DNA material for other future analysis and surveys.

Chapter 4

In this chapter we evaluated and finally confirmed our hypothesis about of the co-spatiality between soil microbial patterns and soil features. This association was confirmed for the three microbial kingdoms although with a high degree of uncertainty in bacterial community whose patterns were more complex.

Our hypothesis was confirmed by analysing both (i) observed and (ii) predicted soil microbial α-diversity patterns in relation with abiotic factors. In the observed study area, we found microbial diversity ranging ~5,794 - 17,323 (Chao1) whereas modelling predictions for the whole of NSW ranged ~ 8,330 - 16,931 (Chao1) for the full community species.

Prior to analysis of the microbial spatial patterns we tested the quality of our observed dataset (i.e. 11,557,499 seqs; 423,740 OTUs) by characterising the microbial communities found in the study area. Therefore, at the beginning of this chapter we described their taxonomy, representativeness in our dataset, richness/evenness, rarity/commonness and diversity distributions across the 49 sampling sites and for disturbed and undisturbed ecosystems. In this analysis we observed that our microbial communities were quite well represented following general tendencies found in soil ecosystems (e.g. *Actinobacteria*, *Ascomycota* most abundant phyla; bacteria/fungi relation 2:1; fungi most evenly distributed than bacteria; archaea the smallest

group). The effect of land use for all of these characteristics showed that natural ecosystems were slightly richer than cropping ecosystems but they hosted more microbial abundance. The richness variation seems to be given by the rare community. In additions, fungi parameters were more affected by disturbed conditions such as cropping ecosystems than bacteria and archaea.

In our first approach into analysing spatial pattern of microbial diversity along the transect we found clear tendencies along both longitudinal and latitudinal environmental gradients. These tendencies showed *archaea* diversity increasing towards the north contrasting with the bacterial and fungal pattern, whereas, the three kingdoms were found to decreases their diversity in a west-east direction.

The patterns observed along the transects were later tested and confirmed by linear and non-linear relations in which we observed that microbial diversity was clearly following soil gradients characterising our transects (e.g. clay, pH, TC; Vertosols in the north, forested humid zones). Further modelling with a *Scorpan approach* using observed and predicted data were consistent with the previous results. Indeed, we found that soil microbial diversity followed similar geographical patterns of soil entities by being modelling microbial, soil physicochemical attributes and other environmental data (25 covariates).

The main findings from these models indicate that *fungi* and *archaea* showed more spatial relation with soil gradients than *bacteria*. In fact, *fungi* and *archaea* were quite well predicted by their respective negative and positive relations with cracking clay conditions (*Vertosols*). In our analysis,

fungi communities have been found highly associated with structural soil attributes and so they can be associated more directly with soil structural conditions. An interesting finding was that predictive maps of *fungi* diversity showed different patterns when taken into account the abundance information by Shannon index. This result together with those found in linear correlations, suggest that richness, diversity and abundance have different patterns and they might be responding to different features and/or soil attributes.

Finally, on the basis of the vast evidence found in the literature, linear and non-linear relations analysis, and modelling focused on digital soil global distribution prediction, our findings suggested that there is a co-spatial pattern between microbial α-diversity and soil entities gradients.

FINAL CONCLUSIONS

- Soil microbial alpha diversity was found to follow geographical patterns across NSW and these patterns were closely associated with the spatial distributions of New South Wales soil types (ASC). Fungi and archaea, more than bacteria, showed co-spatial relation with soil distributions.

- Both microbial patterns, diversity and abundance, seemed to be structured by different abiotic factors.

- Fungi are the microbial kingdom more affected by both soil attributes and land uses.

- Microbial communities were more abundant in cropping ecosystem but more diverse in natural areas.

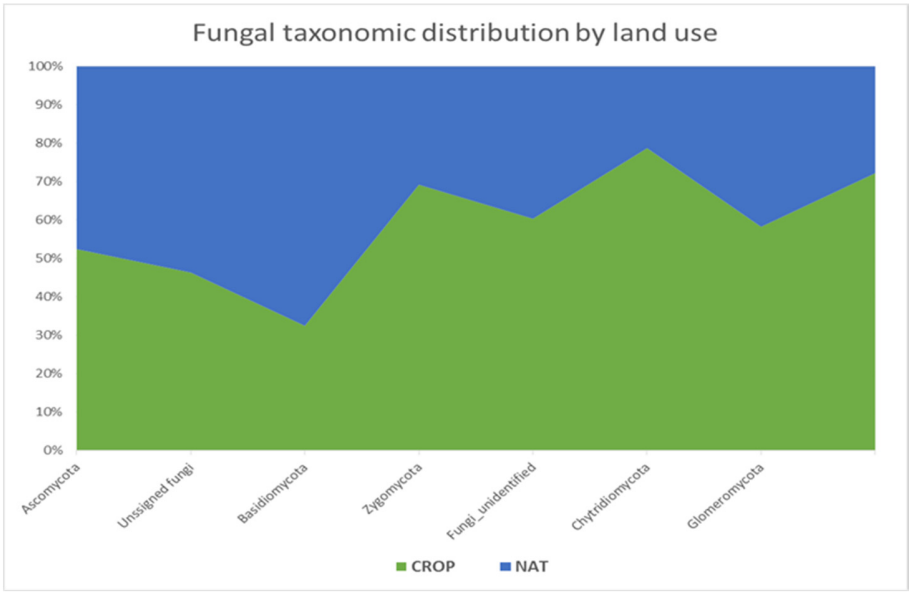
- In the three kingdoms, the rare members of the communities were responsible for increases in diversity.

FUTURE WORK

As presented at beginning of Chapter 1, this investigation was framed to evaluate the complex dimension of biodiversity-pedodiversity relationship. Towards this direction, and as part of a major challenge, we already have evidence of this relationship in our study area at the scale of microbial α-diversities, microbial abundances, taxonomic compositions, and dominance/rarity patterns. In this sense, we have partially confirmed our hypothesis about the biodiversity and pedodiversity spatial associations. Our next challenge will be to move forwards this analysis to the scale of β-diversity and evaluate at what degree of dissimilarities these biotic-abiotic components are related across the space.

APPENDIX 1

Microbial taxonomic distribution affected by land use.



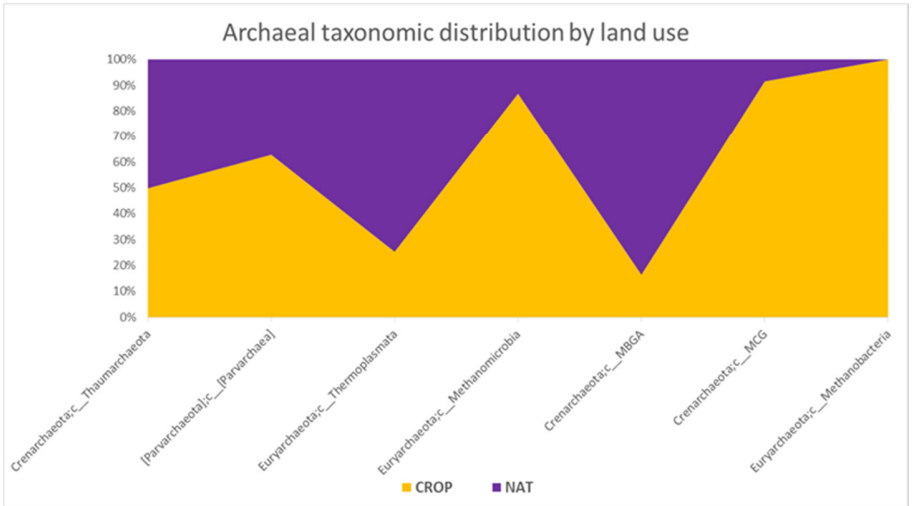


Figure 5−1. Fungal and archaeal taxonomic distribution at phylum level (L2) and class level (L3), respectively.

APPENDIX 2

A. Richness calculations along the two transects in the study area.

Table 5−1. Soil microbial richness per sampling site across $WE_{-transect}$.

#Site	Ecosystem	BACTERIA			FUNGI			ARCHAEA		
		Obs_OTUs	Chao1	PD	Obs_OTUs	Chao1	PD	Obs_OTUs	Chao1	PD
27	CROP	4,849.80	12,599.71	316.47	349.80	452.61	–	13.10	30.14	1.73
	NAT	4,433.50	11,758.86	290.89	445.80	555.92	–	–	–	–
28	CROP	4,321.13	12,894.50	304.04	388.10	526.93	–	–	–	–
	NAT	4,362.90	11,087.04	315.51	539.00	705.55	–	–	–	–
29	CROP	3,362.00	9,215.75	245.68	432.07	665.60	–	–	–	–
	NAT	4,634.10	11,937.20	266.92	517.70	751.64	–	–	–	–
30	CROP	4,176.67	12,340.73	306.73	537.17	729.97	–	–	–	–
	NAT	4,570.75	11,836.70	268.82	254.30	417.06	–	–	–	–
31	CROP	4,564.45	12,368.93	335.90	427.70	650.16	–	–	–	–
	NAT	4,887.85	15,021.39	304.61	402.10	603.43	–	–	–	–
32	CROP	4,813.70	14,772.35	298.78	701.60	992.77	–	–	–	–
	NAT	4,564.23	14,525.54	324.29	622.85	894.76	–	–	–	–
33	CROP	4,723.47	14,721.96	291.23	401.27	642.88	–	7.40	11.80	1.49
	NAT	4,285.07	13,411.69	310.12	623.95	906.54	–	16.10	26.42	2.24
34	CROP	4,114.03	13,570.46	296.49	597.43	892.81	–	–	–	–
	NAT	4,843.95	15,243.39	341.75	559.30	926.15	–	–	–	–
35	CROP	4,449.80	13,187.68	306.41	476.87	713.49	–	–	–	–
	NAT	5,111.10	15,520.41	335.55	604.27	959.13	–	–	–	–
36	CROP	4,031.40	12,467.80	304.98	371.17	551.94	–	20.30	34.67	1.88
	NAT	5,287.00	14,255.58	337.08	466.33	694.96	–	–	–	–
37	CROP	4,484.63	13,814.16	321.81	388.77	564.09	–	–	–	–
	NAT	4,595.40	14,889.47	314.40	500.83	776.89	–	19.40	25.09	1.87
38	CROP	4,549.90	13,377.92	309.41	285.90	380.66	–	24.90	38.19	3.02
	NAT	4,402.97	15,134.45	323.90	500.50	744.59	–	23.90	41.38	2.69
40	CROP	4,517.73	14,253.56	303.25	322.63	496.34	–	23.40	44.37	2.16
	NAT	4,537.77	14,960.86	339.09	451.55	651.86	–	24.90	51.81	2.02
41	CROP	4,247.00	14,078.51	307.82	368.37	566.61	–	19.50	27.12	2.52
	NAT	4,798.57	15,263.16	339.05	422.30	646.19	–	29.25	54.65	2.85
42	CROP	4,289.70	13,741.05	322.26	360.87	522.47	–	20.83	35.66	2.67
	NAT	4,973.37	15,439.49	335.94	392.17	633.02	–	27.00	43.01	2.61
43	CROP	3,714.25	11,320.88	280.22	392.83	573.99	–	–	–	–
	NAT	4,935.47	16,062.00	319.40	613.00	954.14	–	–	–	–
44	CROP	4,589.93	12,518.31	319.31	437.83	592.34	–	24.85	47.04	3.29
	NAT	4,554.87	13,329.46	314.37	464.90	678.40	–	27.30	53.05	2.52
45	CROP	5,458.65	17,630.71	344.49	572.40	907.73	–	–	–	–
	NAT	5,011.93	16,358.77	368.38	575.67	938.97	–	11.50	27.47	2.28
46	CROP	2,702.30	5,796.20	165.14	–	–	–	–	–	–
	NAT	4,489.73	14,798.02	322.73	784.60	1,164.97	–	15.70	42.87	2.29
47	CROP	4,097.45	10,314.46	293.52	–	–	–	21.30	26.94	4.00
	NAT	4,212.93	12,475.46	308.64	564.00	844.39	–	6.50	13.95	1.95
48	CROP	3,534.95	9,674.99	254.27	320.87	374.35	–	14.95	31.15	2.19
	NAT	4,927.70	15,226.18	345.40	543.43	832.01	–	23.25	44.84	2.82

Values are mean of 10 iterations from the rarefied data.

Table 5-2. Soil microbial richness per sampling site across $NS_{-transect}$.

#Site	Ecosystem	BACTERIA			FUNGI			ARCHAEA		
		Obs_OTUs	Chao1	PD	Obs_OTUs	Chao1	PD	Obs_OTUs	Chao1	PD
0	CROP	2,846.27	6,798.92	229.85	222.65	294.55	–	15.17	23.52	2.48
	NAT	–	–	–	349.73	469.86	–	15.00	21.75	2.01
1	CROP	–	–	–	375.27	549.85	–	16.00	30.93	2.26
	NAT	–	–	–	454.73	667.33	–	–	–	–
2	CROP	2,941.30	7,024.83	240.48	395.70	573.17	–	–	–	–
	NAT	3,117.00	7,558.85	258.07	494.10	713.36	–	–	–	–
3	CROP	3,702.47	11,059.77	287.60	433.15	675.86	–	23.20	43.05	2.59
	NAT	2,221.20	4,005.39	185.49	198.60	227.51	–	16.55	26.99	2.99
4	CROP	3,316.70	8,622.10	251.95	534.03	783.59	–	–	–	–
	NAT	3,187.35	9,005.81	255.33	696.50	1,079.15	–	–	–	–
5	CROP	3,244.63	7,427.90	264.83	314.80	371.44	–	16.85	32.04	3.08
	NAT	3,962.33	12,208.02	292.71	649.20	959.46	–	–	–	–
6	CROP	3,320.95	9,237.02	236.28	450.65	657.32	–	15.80	24.43	1.58
	NAT	3,336.20	10,003.12	261.88	538.60	863.19	–	–	–	–
7	CROP	3,763.20	11,268.71	279.60	367.40	569.02	–	–	–	–
	NAT	3,318.90	9,683.18	252.32	791.40	1,123.40	–	9.90	11.87	1.65
8	CROP	3,428.00	9,969.38	266.58	436.30	649.86	–	–	–	–
	NAT	3,041.03	8,683.84	247.28	419.63	643.12	–	10.97	14.56	1.96
9	CROP	4,067.00	13,788.66	306.50	632.27	906.83	–	20.30	43.64	3.40
	NAT	3,594.70	11,345.64	282.00	614.43	942.43	–	10.85	15.10	1.89
10	CROP	3,436.05	10,066.70	275.20	604.90	899.83	–	10.60	27.03	2.72
	NAT	2,737.53	7,058.54	217.72	583.40	857.00	–	–	–	–
11	CROP	3,038.50	8,740.57	247.35	620.47	912.59	–	–	–	–
	NAT	3,542.95	10,781.80	254.83	678.80	1,048.40	–	–	–	–
12	CROP	3,265.87	9,393.04	269.89	509.00	652.29	–	7.73	11.40	1.70
	NAT	3,653.70	12,007.50	256.28	784.87	1,196.13	–	–	–	–
13	CROP	3,430.75	10,637.20	281.24	242.50	401.17	–	9.45	14.73	1.70
	NAT	3,868.37	12,814.47	286.53	685.60	1,061.02	–	12.60	24.95	2.54
14	CROP	3,422.00	9,730.71	278.26	624.50	883.97	–	8.40	11.88	2.24
	NAT	4,033.95	12,133.35	310.38	799.40	1,133.91	–	–	–	–
15	CROP	4,157.95	12,948.22	329.56	394.80	608.33	–	9.80	16.67	2.19
	NAT	3,921.30	10,887.35	301.70	549.67	805.56	–	–	–	–
16	CROP	4,031.70	13,334.71	301.00	595.67	852.75	–	–	–	–
	NAT	2,925.30	7,713.19	210.55	560.10	797.72	–	–	–	–
17	CROP	3,899.57	12,663.38	297.75	451.93	653.71	–	9.40	33.95	2.44
	NAT	3,650.10	11,418.56	283.18	665.87	963.77	–	11.50	16.00	1.67
18	CROP	3,477.80	10,639.72	280.82	483.55	717.72	–	8.00	18.18	1.99
	NAT	–	–	–	674.25	1,009.87	–	–	–	–
19	CROP	3,698.00	10,831.40	297.44	545.10	822.41	–	–	–	–
	NAT	3,231.10	8,600.33	191.74	750.87	1,074.07	–	–	–	–
20	CROP	3,786.57	10,441.13	303.70	489.90	715.02	–	7.10	10.98	2.06
	NAT	3,472.95	10,223.18	270.09	497.40	691.45	–	–	–	–
21	CROP	4,172.43	12,475.13	308.48	666.57	986.17	–	–	–	–
	NAT	3,738.20	10,594.86	285.89	632.13	835.19	–	12.90	20.80	1.76
22	CROP	4,260.23	12,648.40	315.83	399.77	588.04	–	–	–	–
	NAT	3,427.03	9,670.75	274.38	493.70	691.48	–	5.70	5.93	1.42
23	CROP	3,192.77	9,782.88	242.75	539.07	787.99	–	–	–	–
	NAT	3,545.10	10,306.04	273.44	650.37	888.50	–	–	–	–
24	CROP	3,872.10	11,756.61	303.91	565.13	783.88	–	–	–	–
	NAT	3,897.63	11,217.15	310.03	568.50	809.12	–	–	–	–
25	CROP	3,913.57	10,892.30	313.67	617.20	888.86	–	6.30	9.94	1.97
	NAT	4,043.25	12,255.54	286.46	570.73	847.14	–	–	–	–
26	CROP	3,866.65	11,099.98	292.50	597.50	895.67	–	–	–	–
	NAT	3,791.80	10,257.24	296.84	534.45	773.07	–	–	–	–

Values are mean of 10 iterations from the rarefied data.

B. Evenness calculations along the two transects in the study area.

Table 5−3. Soil microbial evenness per sampling site across $WE_{-transect}$.

#Site	Ecosystem	BACTERIA				FUNGI				ARCHAEA			
		H'	1/D	1−D	$E_{1/D}$	H'	1/D	1−D	$E_{1/D}$	H'	1/D	1−D	$E_{1/D}$
27	CROP	11.37	849.99	1.00	0.18	4.92	11.86	0.92	0.03	1.78	1.98	0.50	0.15
	NAT	10.91	339.50	1.00	0.08	5.69	21.68	0.95	0.05	−	−	−	−
28	CROP	10.85	486.80	1.00	0.11	4.32	4.51	0.78	0.01	−	−	−	−
	NAT	11.02	630.24	1.00	0.14	5.72	11.30	0.91	0.02	−	−	−	−
29	CROP	9.42	49.93	0.98	0.01	5.63	19.12	0.95	0.04	−	−	−	−
	NAT	11.11	526.27	1.00	0.11	5.92	17.34	0.94	0.03	−	−	−	−
30	CROP	10.49	174.78	0.99	0.04	6.86	47.13	0.98	0.09	−	−	−	−
	NAT	11.07	442.33	1.00	0.10	3.27	3.92	0.75	0.02	−	−	−	−
31	CROP	11.05	545.83	1.00	0.12	5.34	12.25	0.92	0.03	−	−	−	−
	NAT	11.11	335.50	1.00	0.07	4.75	9.00	0.89	0.02	−	−	−	−
32	CROP	11.11	467.30	1.00	0.10	6.62	21.02	0.95	0.03	−	−	−	−
	NAT	10.67	116.98	0.99	0.03	6.57	32.01	0.97	0.05	−	−	−	−
33	CROP	10.82	230.24	1.00	0.05	4.52	6.63	0.85	0.02	1.35	1.75	0.43	0.24
	NAT	10.40	64.16	0.98	0.01	6.11	13.02	0.92	0.02	2.61	3.25	0.69	0.20
34	CROP	10.00	66.32	0.98	0.02	6.22	19.25	0.95	0.03	−	−	−	−
	NAT	11.07	321.71	1.00	0.07	4.97	7.36	0.86	0.01	−	−	−	−
35	CROP	10.81	342.52	1.00	0.08	5.75	19.24	0.95	0.04	−	−	−	−
	NAT	11.38	815.08	1.00	0.16	5.86	16.12	0.94	0.03	−	−	−	−
36	CROP	10.19	109.14	0.99	0.03	4.85	7.96	0.87	0.02	3.20	5.51	0.82	0.27
	NAT	11.61	1176.89	1.00	0.22	5.69	16.05	0.94	0.03	−	−	−	−
37	CROP	10.77	291.46	1.00	0.06	5.31	12.80	0.92	0.03	−	−	−	−
	NAT	10.71	162.23	0.99	0.04	5.68	13.53	0.93	0.03	3.56	7.71	0.87	0.40
38	CROP	10.83	171.49	0.99	0.04	4.74	9.51	0.89	0.03	4.02	11.83	0.92	0.48
	NAT	10.37	98.05	0.99	0.02	5.88	14.18	0.93	0.03	3.80	9.28	0.89	0.39
39	CROP	9.73	73.73	0.99	0.02	5.53	14.65	0.93	0.04	3.21	5.95	0.83	0.33
	NAT	10.79	287.84	1.00	0.06	5.33	15.76	0.94	0.05	−	−	−	−
40	CROP	10.70	134.07	0.99	0.03	4.12	5.51	0.82	0.02	3.66	7.65	0.87	0.33
	NAT	10.75	230.03	1.00	0.05	5.54	14.25	0.93	0.03	3.90	9.97	0.90	0.40
41	CROP	10.34	108.96	0.99	0.03	4.61	4.37	0.77	0.01	3.60	8.86	0.89	0.45
	NAT	11.05	369.88	1.00	0.08	5.18	7.23	0.86	0.02	4.26	13.19	0.92	0.45
42	CROP	10.63	241.67	1.00	0.06	4.87	9.85	0.90	0.03	3.48	7.10	0.86	0.34
	NAT	11.16	384.67	1.00	0.08	5.18	12.35	0.92	0.03	4.04	10.74	0.91	0.40
43	CROP	9.91	80.90	0.99	0.02	5.21	11.82	0.92	0.03	−	−	−	−
	NAT	11.13	365.09	1.00	0.07	6.44	28.11	0.96	0.05	−	−	−	−
44	CROP	11.08	625.12	1.00	0.14	6.23	25.71	0.96	0.06	3.59	6.04	0.83	0.24
	NAT	10.90	279.54	1.00	0.06	5.22	7.97	0.87	0.02	4.04	10.73	0.91	0.39
45	CROP	11.61	935.47	1.00	0.17	5.90	12.83	0.92	0.02	−	−	−	−
	NAT	11.25	620.76	1.00	0.12	4.95	4.56	0.78	0.01	1.33	1.53	0.35	0.13
46	CROP	9.58	137.79	0.99	0.05	−	−	−	−	−	−	−	−
	NAT	10.69	273.88	1.00	0.06	7.14	40.11	0.98	0.05	2.3	2.66	0.62	0.17
47	CROP	10.82	584.45	1.00	0.14	−	−	−	−	3.66	7.38	0.86	0.35
	NAT	10.66	348.84	1.00	0.08	6.14	21.21	0.95	0.04	0.64	1.20	0.17	0.18
48	CROP	10.10	190.41	0.99	0.05	6.33	31.95	0.97	0.10	2.45	3.17	0.68	0.21
	NAT	11.21	611.13	1.00	0.12	5.23	8.03	0.88	0.01	3.49	5.73	0.83	0.25

Values are mean of 10 iterations from the rarefied data.

Table 5-4. Soil microbial evenness per sampling site across $NS_{-transect}$.

#Site	Ecosystem	BACTERIA				FUNGI				ARCHAEA			
		H'	1/D	1-D	$E_{1/D}$	H'	1/D	1-D	$E_{1/D}$	H'	1/D	1-D	$E_{1/D}$
0	CROP	9.34	73.57	0.99	0.03	5.15	15.00	0.93	0.07	2.80	4.52	0.78	0.30
	NAT	–	–	–	–	5.11	7.69	0.87	0.02	3.02	5.55	0.82	0.37
1	CROP	–	–	–	–	4.77	7.18	0.86	0.02	2.83	4.59	0.78	0.29
	NAT	–	–	–	–	5.78	19.29	0.95	0.04	–	–	–	–
2	CROP	9.46	88.26	0.99	0.03	5.46	17.38	0.94	0.04	–	–	–	–
	NAT	9.98	225.95	1.00	0.07	6.08	23.21	0.96	0.05	–	–	–	–
3	CROP	9.99	186.67	0.99	0.05	5.65	19.34	0.95	0.04	3.7	9.09	0.89	0.39
	NAT	9.33	184.07	0.99	0.08	5.68	27.33	0.96	0.14	2.99	5.27	0.81	0.32
4	CROP	9.86	127.19	0.99	0.04	6.07	20.81	0.95	0.04	–	–	–	–
	NAT	9.51	120.01	0.99	0.04	6.69	34.05	0.97	0.05	–	–	–	–
5	CROP	9.91	152.00	0.99	0.05	5.31	13.98	0.93	0.04	2.99	5.40	0.81	0.32
	NAT	10.42	294.84	1.00	0.07	6.91	46.73	0.98	0.07	–	–	–	–
6	CROP	10.11	286.97	1.00	0.09	6.12	28.48	0.96	0.06	2.72	3.72	0.73	0.24
	NAT	9.96	165.58	0.99	0.05	5.31	8.70	0.89	0.02	–	–	–	–
7	CROP	10.3	245.54	1.00	0.07	4.96	12.44	0.92	0.03	–	–	–	–
	NAT	9.86	185.44	0.99	0.06	7.23	52.87	0.98	0.07	2.39	3.79	0.74	0.38
8	CROP	10.18	310.85	1.00	0.09	5.73	18.68	0.95	0.04	–	–	–	–
	NAT	9.71	208.06	0.99	0.07	4.32	4.63	0.78	0.01	1.92	2.26	0.56	0.21
9	CROP	10.67	435.02	1.00	0.11	6.78	36.36	0.97	0.06	3.17	5.58	0.82	0.27
	NAT	9.92	175.26	0.99	0.05	6.00	10.09	0.90	0.02	2.25	3.23	0.69	0.30
10	CROP	10.3	419.61	1.00	0.12	6.69	36.04	0.97	0.06	1.09	1.38	0.28	0.13
	NAT	9.32	126.92	0.99	0.05	5.26	6.23	0.84	0.01	–	–	–	–
11	CROP	9.91	262.43	1.00	0.09	6.83	38.37	0.97	0.06	–	–	–	–
	NAT	10.03	168.72	0.99	0.05	6.02	9.46	0.89	0.01	–	–	–	–
12	CROP	10.14	347.55	1.00	0.11	6.41	28.97	0.97	0.06	1.64	2.28	0.56	0.30
	NAT	10.10	243.29	1.00	0.07	6.91	31.05	0.97	0.04	–	–	–	–
13	CROP	10.09	253.75	1.00	0.07	2.96	3.52	0.72	0.01	2.09	2.86	0.65	0.30
	NAT	10.51	420.72	1.00	0.11	6.68	30.23	0.97	0.04	2.83	5.30	0.81	0.42
14	CROP	10.34	410.79	1.00	0.12	6.65	25.13	0.96	0.04	1.61	2.12	0.53	0.25
	NAT	10.75	557.58	1.00	0.14	7.01	33.10	0.97	0.04	–	–	–	–
15	CROP	10.77	503.84	1.00	0.12	4.52	4.34	0.77	0.01	1.37	1.60	0.37	0.16
	NAT	10.74	591.28	1.00	0.15	6.08	20.03	0.95	0.04	–	–	–	–
16	CROP	10.63	441.57	1.00	0.11	6.47	28.51	0.96	0.05	–	–	–	–
	NAT	9.54	181.04	0.99	0.06	6.13	17.85	0.94	0.03	–	–	–	–
17	CROP	10.52	363.65	1.00	0.09	5.75	15.77	0.94	0.03	0.93	1.32	0.24	0.14
	NAT	10.21	262.04	0.99	0.07	6.82	37.44	0.97	0.06	2.66	4.82	0.79	0.42
18	CROP	10.23	296.45	1.00	0.09	6.19	23.79	0.96	0.05	0.8	1.26	0.21	0.16
	NAT	–	–	–	–	6.10	15.67	0.94	0.02	–	–	–	–
19	CROP	10.34	330.32	1.00	0.09	6.50	32.96	0.97	0.06	–	–	–	–
	NAT	9.41	118.69	0.99	0.04	6.91	30.14	0.97	0.04	–	–	–	–
20	CROP	10.24	357.79	0.99	0.09	5.85	15.86	0.94	0.03	1.11	1.49	0.33	0.21
	NAT	10.35	403.34	1.00	0.12	6.19	26.47	0.96	0.05	–	–	–	–
21	CROP	10.85	587.11	1.00	0.14	6.85	35.36	0.97	0.05	–	–	–	–
	NAT	10.54	434.78	1.00	0.12	6.97	50.24	0.98	0.08	2.49	3.56	0.72	0.28
22	CROP	10.84	473.99	1.00	0.11	5.38	12.41	0.92	0.03	–	–	–	–
	NAT	10.21	302.92	1.00	0.09	5.95	13.11	0.92	0.03	1.27	1.69	0.41	0.30
23	CROP	9.62	139.53	0.99	0.04	6.30	20.03	0.95	0.04	–	–	–	–
	NAT	10.19	287.07	1.00	0.08	7.19	60.51	0.98	0.09	–	–	–	–
24	CROP	10.65	502.19	1.00	0.13	6.62	29.06	0.97	0.05	–	–	–	–
	NAT	10.42	280.19	1.00	0.07	6.57	33.39	0.97	0.06	–	–	–	–
25	CROP	10.69	504.53	1.00	0.13	6.69	31.67	0.97	0.05	0.7	1.23	0.19	0.20
	NAT	10.59	376.8	1.00	0.09	4.71	3.07	0.67	0.01	–	–	–	–
26	CROP	10.67	536.14	1.00	0.14	6.29	24.01	0.96	0.04	–	–	–	–
	NAT	10.38	299.07	1.00	0.08	5.82	14.57	0.93	0.03	–	–	–	–

Values are mean of 10 iterations from the rarefied data.

APPENDIX 3

Correlations found between microbial and abiotic parameters.

Table 5–5. Pearson correlation (r) between microbial abundances and abiotic parameters.

	BACTERIA							ARCHAEA							FUNGI				
	Actino-bacteria	*Proteo-bacteria*	*Acido-bacteria*	*Firmicutes*	*Verruco-microbia*	*Planctomycetes*	*Cyano-bacteria*	*Thaumar-chaeota*	*Parvar-chaea*	*Thermo-plasmata*	*Methano-microbia*	*MBGA*	*MCG*	*Methano-bacteria*	*Asco-mycota*	*Basidio-mycota*	*Zygo-mycota*	*Glomero-mycota*	*Chytridio-mycota*
Soil physicochemical properties																			
pH ($CaCl_2$)	**0.45**	0.01	**-0.51**	**-0.17**	**-0.27**	**-0.42**	**0.14**	-0.10	0.05	0.19	0.05	-0.22	-0.02	-0.02	0.06	**-0.27**	0.11	-0.12	**0.24**
pH (H_2O)	**0.45**	-0.06	**-0.47**	**-0.19**	**-0.30**	**-0.43**	**0.16**	-0.15	0.05	0.22	0.09	-0.20	0.02	0.02	0.06	**-0.28**	**0.13**	**-0.13**	**0.22**
TC	-0.08	**0.34**	0.01	0.09	**0.23**	0.08	**-0.20**	0.18	-0.12	-0.06	-0.16	-0.03	-0.12	-0.13	**-0.13**	**0.29**	**-0.15**	0.04	-0.11
TN	-0.09	**0.32**	-0.05	**0.15**	**0.22**	0.07	**-0.21**	0.19	-0.14	-0.06	-0.15	-0.07	-0.12	-0.12	-0.08	**0.25**	**-0.14**	0.06	-0.10
P	-0.04	**0.17**	**-0.30**	**0.29**	-0.09	**-0.18**	-0.12	0.13	-0.19	0.01	-0.08	-0.08	-0.08	-0.08	0.08	-0.01	-0.10	-0.08	0.02
K	**0.32**	**0.15**	**-0.46**	-0.08	**-0.23**	**-0.30**	-0.08	0.21	-0.21	0.01	-0.14	**-0.25**	-0.17	-0.17	0.07	-0.11	0.03	**-0.15**	0.06
NH_4-N	**-0.17**	**0.22**	0.01	**0.25**	0.12	-0.02	**-0.18**	-0.09	-0.09	**0.39**	-0.12	-0.04	-0.09	-0.09	-0.05	**0.16**	-0.12	0.00	-0.06
NO_3-N	-0.04	**0.20**	**-0.23**	0.11	-0.01	0.00	-0.11	0.18	-0.21	-0.08	-0.05	-0.10	-0.08	-0.08	-0.05	0.00	0.02	-0.03	0.03
EC	**0.15**	0.24	**-0.36**	-0.02	-0.01	0.05	**-0.15**	0.07	-0.17	**0.23**	-0.12	-0.18	-0.15	-0.15	-0.10	-0.01	0.10	0.06	0.09
ECEC	**0.44**	**0.16**	**-0.48**	**-0.16**	**-0.14**	**-0.35**	-0.06	0.08	-0.11	0.16	-0.10	-0.21	-0.17	-0.17	0.04	-0.09	0.05	**-0.15**	0.08
Exc. Al	**-0.27**	0.00	**0.48**	-0.10	0.07	0.17	-0.09	-0.04	-0.22	0.05	0.11	**0.31**	0.09	0.09	-0.10	**0.22**	-0.08	0.00	-0.09
Exc. Ca	**0.42**	**0.19**	**-0.51**	**-0.13**	-0.09	**-0.31**	-0.06	0.09	-0.09	0.15	-0.11	-0.21	-0.17	-0.17	0.01	-0.06	0.08	-0.12	0.09
Exc. Mg	**0.43**	0.09	**-0.39**	**-0.18**	**-0.17**	**-0.39**	-0.06	0.06	-0.11	0.13	-0.05	-0.17	-0.13	-0.13	0.10	**-0.14**	-0.03	**-0.18**	0.05
Exc. K	**0.33**	0.12	**-0.43**	-0.12	**-0.23**	**-0.26**	-0.06	0.21	-0.21	0.03	-0.16	**-0.26**	-0.19	-0.19	0.06	**-0.13**	0.06	**-0.14**	0.07
Exc. Na	**0.37**	-0.01	**-0.28**	**-0.17**	**-0.21**	**-0.30**	-0.04	**-0.23**	-0.03	**0.47**	0.05	-0.10	-0.02	-0.02	0.05	**-0.20**	-0.01	-0.08	0.04
Clay	**0.27**	0.04	**-0.27**	-0.08	0.00	**-0.35**	-0.05	0.08	-0.11	0.15	-0.11	-0.11	-0.17	-0.18	**0.13**	**-0.16**	0.07	**-0.20**	0.08
T-Hor	-0.10	0.00	0.07	0.12	0.12	0.01	-0.08	-0.06	-0.17	-0.10	0.20	0.01	**0.25**	**0.24**	0.04	0.08	-0.11	-0.04	-0.06
N-Hor	0.08	**-0.16**	0.08	-0.06	-0.02	**0.20**	0.03	-0.17	-0.06	0.09	0.17	0.02	0.22	0.22	-0.05	0.02	0.07	**0.28**	0.01
SI	**0.19**	0.07	**-0.27**	-0.06	**-0.15**	**-0.48**	0.02	-0.11	-0.13	0.02	**0.24**	-0.09	0.18	0.18	**0.15**	**-0.21**	0.04	**-0.34**	**0.16**
Wetness index	**-0.23**	**0.22**	**0.16**	**0.14**	**0.23**	**0.17**	**-0.25**	0.18	-0.15	0.02	-0.19	**0.27**	**-0.22**	**-0.23**	**0.17**	0.01	0.00	0.06	-0.06
Other environmental attributes																			
Landsat B1	0.01	0.05	-0.08	**0.17**	**-0.19**	**-0.16**	-0.06	0.13	-0.14	-0.04	-0.05	-0.15	-0.09	-0.09	0.12	-0.11	-0.01	-0.09	0.04
Landsat B2	0.03	0.08	-0.12	**0.17**	**-0.19**	**-0.25**	-0.05	0.11	-0.13	-0.06	-0.01	-0.17	-0.04	-0.04	0.11	**-0.13**	-0.04	**-0.13**	0.07
Landsat B3	0.08	0.04	**-0.17**	0.13	-0.28	**-0.35**	0.02	0.01	-0.09	-0.08	0.11	-0.19	0.09	0.09	0.09	**-0.16**	-0.05	**-0.19**	0.10
Landsat B4	**-0.17**	**0.33**	**-0.14**	**0.30**	0.08	-0.12	**-0.19**	0.12	-0.16	-0.13	0.01	-0.11	0.03	0.03	0.05	0.04	**-0.14**	-0.06	-0.01
Landsat B5	0.07	0.07	**-0.26**	**0.14**	**-0.23**	**-0.42**	0.10	-0.03	0.01	-0.13	0.15	-0.22	0.16	0.16	-0.05	-0.08	-0.06	**-0.24**	0.11
Landsat B7	**0.14**	-0.01	**-0.23**	0.03	**-0.30**	**-0.42**	**0.14**	-0.09	-0.01	-0.08	0.20	-0.22	0.20	0.20	0.00	**-0.17**	-0.01	**-0.26**	0.12
Elevation	**-0.42**	**0.26**	**0.18**	**0.36**	**0.57**	0.07	**-0.16**	0.13	0.00	-0.07	-0.16	0.07	-0.12	-0.12	0.00	**0.16**	-0.11	0.00	-0.11
FitDose (U, Th, K)	**-0.22**	**0.18**	0.01	**0.17**	**0.43**	**0.26**	**-0.14**	0.19	-0.11	0.00	-0.19	-0.12	-0.19	-0.19	0.01	0.05	0.05	0.04	-0.02
NDVI	**-0.27**	**0.30**	-0.01	**0.20**	**0.35**	0.07	**-0.15**	-0.01	-0.04	-0.12	0.08	0.04	0.14	0.14	-0.10	**0.22**	**-0.17**	0.06	-0.09
NDWI	**-0.23**	**0.23**	0.11	**0.16**	**0.29**	**0.27**	**-0.23**	0.14	-0.13	0.01	-0.16	0.15	-0.15	-0.15	0.09	**0.14**	-0.08	**0.20**	-0.11
Precipitation	-0.11	**0.29**	**-0.14**	**0.18**	**0.18**	-0.10	**-0.18**	0.10	-0.12	0.07	-0.11	0.06	-0.14	-0.14	-0.03	0.10	-0.09	-0.04	-0.07
Slope	**-0.26**	0.12	**0.18**	**0.19**	0.11	**0.18**	-0.13	0.12	-0.07	-0.07	-0.09	-0.07	-0.06	-0.06	**-0.13**	**0.24**	**-0.13**	0.08	-0.07
Land surface T°	**0.45**	-0.11	**-0.37**	**-0.19**	**-0.37**	**-0.52**	**0.16**	-0.11	0.02	0.00	0.17	-0.20	0.14	0.14	-0.03	**-0.15**	-0.03	**-0.23**	**0.13**
Wetness index	**-0.23**	**0.22**	**0.16**	**0.14**	**0.23**	**0.17**	**-0.25**	0.18	-0.15	0.02	-0.19	**0.27**	**-0.22**	**-0.23**	**0.17**	0.01	0.00	0.06	-0.06

*Significant correlations are indicated in bold (P≤0.05) in accordance with a Pearson's paired sample association test.

Table 5–6. Pearson correlation (r) between microbial diversity and abiotic parameters.

	BACTERIA				ARCHAEA				FUNGI		
	Chao1	PD	H'	I-D	Chao1	PD	H'	1-D	Chao1	H'	1-D
Soil physicochemical properties											
pH ($CaCl_2$)	**0.32**	**0.40**	**0.18**	-0.12	**0.63**	**0.53**	**0.59**	**0.48**	**-0.41**	**-0.32**	**-0.17**
pH (H_2O)	**0.26**	**0.31**	0.11	**-0.15**	**0.64**	**0.57**	**0.66**	**0.57**	**-0.41**	**-0.30**	**-0.14**
TC	0.04	0.01	**0.24**	0.12	**-0.25**	**-0.35**	**-0.34**	**-0.35**	0.12	-0.07	**-0.22**
TN	0.09	0.08	**0.27**	0.12	-0.21	**-0.34**	**-0.33**	**-0.36**	0.09	-0.09	**-0.24**
P	**0.14**	**0.17**	**0.23**	-0.03	0.01	**-0.23**	-0.06	-0.08	**-0.13**	**-0.13**	-0.03
K	**0.38**	**0.39**	**0.31**	0.01	**0.43**	0.02	**0.37**	**0.31**	**-0.26**	**-0.29**	**-0.22**
NH_4-N	-0.02	-0.05	0.12	0.03	**-0.34**	**-0.25**	**-0.27**	-0.22	-0.02	-0.02	-0.05
NO_3-N	0.04	0.09	**0.18**	0.07	-0.11	**-0.30**	-0.15	-0.15	**-0.25**	**-0.25**	**-0.18**
EC	0.09	**0.17**	**0.23**	0.13	-0.07	-0.17	-0.07	-0.08	**-0.24**	**-0.27**	**-0.24**
ECEC	**0.25**	**0.27**	**0.18**	**-0.17**	**0.44**	**0.26**	**0.57**	**0.54**	**-0.39**	**-0.39**	-0.29
Exc. Al	**-0.28**	**-0.43**	**-0.28**	-0.03	-0.13	-0.15	0.09	0.15	**0.17**	**0.16**	0.07
Exc. Ca	**0.27**	**0.30**	**0.21**	**-0.17**	**0.44**	**0.25**	**0.53**	**0.50**	**-0.39**	**-0.43**	**-0.36**
Exc. Mg	**0.19**	**0.18**	0.11	**-0.18**	**0.41**	**0.26**	**0.59**	**0.58**	**-0.34**	**-0.28**	**-0.13**
Exc. K	**0.33**	**0.35**	**0.24**	-0.01	**0.41**	0.03	**0.36**	**0.30**	**-0.27**	**-0.27**	**-0.22**
Exc. Na	0.11	0.09	0.01	-0.12	0.13	**0.31**	**0.33**	**0.36**	**-0.28**	**-0.18**	-0.01
Clay	0.04	-0.03	-0.06	**-0.22**	**0.28**	0.14	**0.48**	**0.53**	**-0.41**	**-0.20**	0.01
T-Hor	-0.03	-0.10	-0.10	-0.07	-0.18	-0.22	-0.20	-0.21	-0.06	-0.11	-0.07
N-Hor	-0.07	-0.03	-0.11	0.05	-0.12	-0.12	-0.10	-0.10	**0.27**	**0.26**	0.12
SI	**0.13**	0.12	0.07	**-0.16**	**0.48**	**0.33**	**0.60**	**0.58**	**-0.48**	**-0.28**	-0.01
Other environmental attributes											
Landsat B1	0.05	-0.01	0.04	**-0.15**	**0.25**	0.12	**0.34**	**0.25**	**-0.15**	-0.08	0.02
Landsat B2	**0.13**	0.05	0.09	**-0.16**	**0.36**	0.16	**0.40**	**0.30**	**-0.19**	**-0.13**	0.00
Landsat B3	**0.23**	**0.14**	**0.14**	-0.12	**0.46**	**0.27**	**0.41**	**0.29**	**-0.20**	**-0.15**	-0.02
Landsat B4	**0.16**	0.12	**0.29**	0.00	-0.01	-0.22	-0.06	-0.12	-0.10	**-0.13**	-0.08
Landsat B5	**0.41**	**0.35**	**0.32**	0.02	**0.39**	**0.24**	0.21	0.07	**-0.15**	**-0.18**	-0.12
Landsat B7	**0.34**	**0.29**	**0.22**	-0.02	**0.41**	**0.33**	**0.31**	0.19	**-0.20**	**-0.14**	-0.04
Elevation	0.03	-0.05	0.10	-0.05	**-0.67**	**-0.46**	**-0.56**	**-0.44**	0.02	0.02	0.02
FitDose (U, Th, K)	0.09	0.10	-0.01	-0.12	**-0.35**	**-0.29**	**-0.24**	**-0.23**	**0.19**	**0.14**	0.06
NDVI	0.11	0.12	**0.27**	**0.17**	**-0.29**	**-0.38**	**-0.42**	**-0.40**	**0.12**	-0.01	-0.10
NDWI	**-0.20**	**-0.18**	-0.02	0.00	**-0.45**	**-0.45**	**-0.32**	-0.21	0.05	0.04	0.03
Precipitation	0.05	0.04	**0.20**	0.01	-0.12	-0.20	0.00	0.08	**-0.17**	**-0.20**	-0.10
Slope	-0.12	**-0.19**	-0.01	-0.03	**-0.55**	**-0.38**	**-0.62**	**-0.64**	**0.13**	0.03	-0.02
Land surface T°	**0.36**	**0.33**	**0.16**	-0.07	**0.69**	**0.51**	**0.63**	**0.49**	**-0.24**	**-0.26**	**-0.15**
Wetness index	**-0.23**	**-0.20**	-0.11	-0.12	**-0.29**	**-0.33**	-0.09	0.01	**-0.14**	-0.02	0.08

*Significant correlations are indicated in bold (P≤0.05) in accordance with a Pearson's paired sample association test.

APPENDIX 4

Sources of the environmental covariates used as input for microbial diversity modelling in NSW, Australia.

Table 5−7. Environmental variables used for modelling microbial diversity.

Layer	Pixel resolution (m)	Product details	Source
Clay, Organic carbon, pH, Silt, Sand, ECEC, Bulk density, Available water capacity, Total P, Total Nitrogen	90	TERN's Soil and Landscape grid of Australia	(Grundy et al., 2015)
Land surface temperature	928	MYD11A1.005 Land Surface Temperature and Emissivity Daily Global 1 km Grid SIN. NASA LP DAAC at the USGS EROS Center.	(Google Earth Engine Team, 2015)
Elevation	30	The Shuttle Radar Topography Mission	(Google Earth Engine Team, 2015)
Slope	30	Derived from The Shuttle Radar Topography Mission	(Google Earth Engine Team, 2015)
NDWI	30	Landsat 7 8-Day NDWI Composite Image.	(Google Earth Engine Team, 2015)
NDVI	30	Derived from USGS Landsat 7 TOA Reflectance (Orthorectified) with mask, USGS.	(Google Earth Engine Team, 2015)
Precipitation	90	2014 Australian national map layers	(ABARES, 2014)
Salinity		2014 Australian national map layers	(ABARES, 2014)
Gamma-radiometrics	100	The Radiometric Map of Australia	(Minty et al., 2009)
Landsat 7-Bands1,2,3,4,5,7	30	Landsat 7 TOA Reflectance (Orthorectified) with mask, USGS.	(Google Earth Engine Team, 2015)
Gap-filled tasseled cap wetness index	5000	MODIS tasselled cap: land cover characteristics expressed through transformed MODIS data.	(Google Earth Engine Team, 2015)