Actinobacterial Diversity in Atacama Desert Habitats as a Road Map to Biodiscovery

A thesis submitted by

Hamidah Idris

for the award of Doctor of Philosophy

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

July 2016

School of Biology,
Faculty of Science, Agriculture and Engineering,

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

Abstract

The Atacama Desert of Northern Chile, the oldest and driest nonpolar desert on the planet, is known to harbour previously undiscovered actinobacterial taxa with the capacity to synthesize novel natural products. In the present study, culture-dependent and culture-independent methods were used to further our understanding of the extent of actinobacterial diversity in Atacama Desert habitats. The culture-dependent studies focused on the selective isolation, screening and dereplication of actinobacteria from high altitude soils from Cerro Chajnantor. Several strains, notably isolates designated H9 and H45, were found to produce new specialized metabolites. Isolate H45 synthesized six novel metabolites, lentzeosides A-F, some of which inhibited HIV-1 integrase activity. Polyphasic taxonomic studies on isolates H45 and H9 showed that they represented new species of the genera *Lentzea* and *Streptomyces,* respectively; it is proposed that these strains be designated as *Lentzea chajnantorensis* sp. nov. and *Streptomyces aridus* sp. nov.. Additional isolates from sampling sites on Cerro Chajnantor were considered to be nuclei of novel species of *Actinomadura, Amycolatopsis, Cryptosporangium* and *Pseudonocardia.*

A majority of the isolates produced bioactive compounds that inhibited the growth of one or more strains from a panel of six wild type microorganisms while those screened against *Bacillus subtilis* reporter strains inhibited sporulation and cell envelope, cell wall, DNA and fatty acid synthesis.

Initial culture-independent studies were carried out to establish the extent of actinobacterial diversity in a range of hyper- and extreme hyper-arid Atacama Desert soils. Community DNA extracted from soil collected from the sampling sites was surveyed for actinobacteria by 454 pyrosequencing; rarefaction analyses indicated good coverage at most of the sites. The results revealed an amazing and unexpected taxonomic diversity at the ranks of order, family and genus, much of it novel. The total number of genera, for instance, is 328, of which around 40% could not be assigned to validly published genera. Rank abundancy profiles indicated that much of this diversity can be attributed to low abundancy taxa. Similar results were obtained from community DNA extracted from surface and subsurface soil samples collected at three different altitudes on Cerro Chajnantor. Actinobacterial community structure at these sampling sites was influenced by altitude and sampling depth, as well as several environmental variables that included conductivity, pH, redox potential and organic matter content.

It is evident from these studies that the Atacama Desert landscape abounds in novel actinobacterial taxa that synthesize a broad range of specialized metabolites that can be developed as drug leads.

Table of Contents

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

Chapter 1. General Introduction

1.1 Background

Seventy years ago Sir Alexander Fleming, who discovered penicillin, warned in his Nobel Prize acceptance speech that the misuse of antibiotics would lead to the emergence of drug resistant microbial pathogens (Fleming 1945). Fleming's prediction has come to pass as antibiotic-resistant microbial pathogens now threaten the foundations of modern medicine. Routine medical procedures such as heart bypass surgery, cancer therapy, trauma surgery and intensive care treatments depend on the effective use of antibiotics. Indeed, some patients, such as those with cystic fibrosis, require regular drug therapy for this and many other infections by antibiotic-resistant pathogens that can be fatal.

The selective pressures on microorganisms led to the emergence and spread of multiple drug resistant (MDR) pathogens, as exemplified by the ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanii, Pseudomonas aeruginosa* and *Enterobacter* species) all of which are now resistant to most antibiotics (Payne et al. 2007; Boucher et al. 2009; Genilloud 2014). MDR Gram-negative bacteria, in particular, are frequently associated with high mortality rates and a lack of treatment options for hospital acquired infections (Silver 2011).This problem extends to foodborne diseases caused by *Clostridium difficile, Escherichia coli* and *Salmonella* species and to pathogens like *Mycobacterium tuberculosis* and *Neisseria gonorrhoeae* (Wright 2012). It seems somewhat counterintuitive that the number of therapeutic drug leads is in sharp decline given the prospect of a return to the pre-antibiotic days of medicine (Talbot et al. 2006; Spellberg et al. 2008; Donadio et al. 2010; Butler and Cooper 2011; Genilloud 2014).

Microbial natural products remain the most promising source of new therapeutic drugs (Baltz 2007, 2008; Cragg and Newman 2013). This is partly because alternative approaches, such as target-based screening of chemical libraries have been found wanting, but also because most antibiotics are produced by microbes that have been evolving for around a billion years hence their fitness has been tested by an ability to pass through outer membranes and inhibit target enzymes, micromolecules and macromolecular structures (Baltz 2008). It is now evident that natural product libraries encompass much greater structural diversity than chemical libraries derived from combinatorial chemistry (Dobson 2004; Drees et al. 2004). More recently vast amounts of genomic data generated from whole-genome sequencing projects can be exploited in

the search for novel drug leads (Corre and Challis 2009; Walsh and Fischbach 2010; Winter et al. 2011).

The ability of microorganisms to synthesize natural products is unevenly distributed across the microbial world. Amongst prokaryotes, organisms classified in the phylum *Actinobacteria* (Goodfellow 2012b) are a rich source of bioactive compounds (Hopwood 2007; Newman and Cragg 2012). Indeed, about 75% of antibiotics are produced by filamentous bacteria belonging to the class *Actinobacteria* (Goodfellow 2012a), though most are made by members of a single taxon, the genus *Streptomyces* (Bérdy 2012; Demain 2014). The genomes of filamentous actinobacteria, unlike those of most prokaryotes, are rich in biosynthetic gene clusters that code for known or predicted specialized metabolites (Goodfellow and Fiedler 2010; Tang et al. 2015), a development that renewed interest in these organisms as a source of new antibiotics (Baltz 2007, 2008; Donadio et al. 2010; Gomez-Escribano and Bibb 2014). Culture-independent surveys make it clear that a vast array of taxonomically novel actinobacteria are present in natural ecosystems (Das et al. 2007; Yang et al. 2012; Serkebaeva et al. 2013; Prieto-Davó et al. 2013), an unseen majority potentially available for bioprospecting campaigns (Whitman et al. 1998; Bull 2004; Genilloud 2014).

Despite the developments outlined above it has become increasingly difficult to discover novel drug leads from filamentous actinobacteria isolated from well trawled terrestrial habitats as screening such organisms leads to the costly rediscovery of known compounds (Baltz 2007; Busti et al. 2006; Genilloud 2014). This problem can be addressed by the application of new search and discovery strategies, as exemplified by the taxonomic approach to bioprospecting recommended by Goodfellow and Fiedler (2010). The key steps involved in this culture-dependent strategy are shown in Figure 1.1. The first step in the antibiotic pipeline involves the choice of environmental samples followed by the selective isolation and recognition of putatively novel filamentous actinobacteria. Subsequent steps include the detection of interesting bioactive compounds from dereplicated strain libraries, primary screening of fermentation broths and mycelial extracts using chemical procedures such as HPLC diode array screening. The final steps are focused on the detection of putatively novel specialized metabolites and their structural elucidation. The premise underlying this taxonomic approach to bioprospecting is that extreme environmental conditions give rise to a unique actinobacterial diversity which is the basis of novel chemistry (Bull and Stach 2007; Bull 2011; Goodfellow et al. 2013).

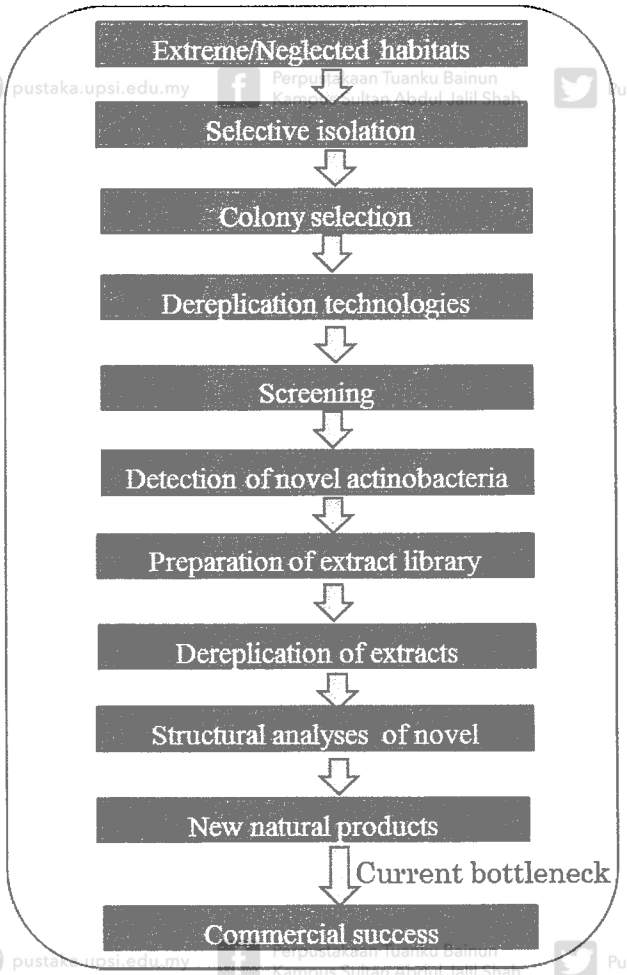

Figure 1.1 Culture-dependent bioprospecting strategy (modified from Goodfellow & Fiedler, 2010).

In practice, novel actinobacteria, especially streptomycetes, have been isolated from marine habitats using the strategy outlined in Figure 1.1 and representatives of dereplicated taxa shown to produce novel antibiotics with unique modes of action (Bull et al. 2005; Fiedler et al. 2005; Goodfellow and Fiedler 2010). Novel actinobacteria from deep sea sediments are a particularly good source of new specialized metabolites, as illustrated by the discovery of a new family of polyketides, the abyssomycins, from *Verrucosispora maris* AB-18-032^T (Riedlinger et al. 2004; Goodfellow et al. 2012b), the anticancer drug, salinosporamide from *Salinispora tropica* CNB-440^T (Jensen et al. 2007; Fenical et al. 2009) and the dermacozines, from *Dermacoccus abyssi* MT1.1^T and MT 1.2, non-filamentous, piezzotolerant strains isolated from the Challenger Deep of the Mariana Trench (Pathom-Aree et al. 2006a; Abdel-Mageed et al. 2010; Wagner et al. 2014). Strong support for culture-dependent approaches to bioprospecting come from extensive surveys of the obligate marine genus *Salinispora* (Jensen et al. 2005; Freel et al. 2012; Ahmed et al. 2013), notably from representative strains which show clear

evidence of coupling between taxonomic and chemical diversity (Jensen 2010; Ziemert et al. 2014).

More recently, the taxonomic approach to the discovery of new specialized metabolites from novel actinobacteria was extended to another extreme biome, the temperate Atacama Desert, the oldest and driest desert on Earth which is located in Northern Chile (Bull and Asenjo 2013; Bull et al. 2016). In initial studies, small but taxonomically diverse populations of presumptively novel filamentous actinobacteria were isolated from hyper-arid and extreme hyper-arid soils collected from the Salar de Atacama and Yungay Core regions of the desert, respectively (Okoro et al. 2009; Busarakam 2014). Subsequently, several new species of *Lechevalieria* and *Streptomyces* were validly published (Okoro et al. 2010; Santhanam et al. 2012b; 2012a; 2013), notably *Streptomyces leeuwenhoekii* (Busarakam et al. 2014), the type strain of this species produces new polyketide antibiotics, the chaxalactins (Rateb et al. 2011b) and chaxamycins (Rateb et al. 2011a) while additional members synthesize 22-membered macrolactone antibiotics, the atacamycins (Nachtigall et al. 2011; Elsayed et al. 2015) and chaxapeptin, a new lasso peptide (Elsayed et al. 2015). Rateb et al. (2013) reported the induction of biosynthetic pathways in an *Aspergillus fumigatus* strain when grown in the presence of *Streptomyces bullii* $C2^T$ leading to the isolation of ergosterol, seven metabolites belonging to the diketopiperazine class of alkaloids and two metabolites from the rare class of pseurotins.

1.2 Objectives and content of thesis

This project was designed to build upon and extend pioneering work on Atacama Desert actinobacteria carried out by Okoro et al. (2009) and Busarakam (2014), the results of which have been considered within the broader context of Atacama Desert microbiology (Bull and Asenjo 2013; Bull et al. 2016). A biphasic approach was adopted in this investigation as environmental samples collected from a range of Atacama Desert habitats were examined using culture-dependent and culture-independent procedures, as outlined in Figure 1.2. The culture-independent studies were intended to establish the extent of actinobacteria diversity in Atacama Desert habitats with an emphasis on determining and explaining actinobacterial community structure in soil samples collected at altitudes between *ca.* 3000 and 5000 m. up to the Chajnantor Plateau at the Atacama Large Millimeter Array (ALMA) Observation site. The culture-dependent studies were focused on the isolation, dereplication and screening of actinobacteria isolated from the ALMA

samples and from extreme hyper-arid soil collected from Lomas Bayas, possibly the driest area in the Atacama Desert. Selected isolates from these and earlier studies were the subject of polyphasic taxonomic analyses while extracts of judiciously chosen isolates showing bioactivity in plug assays were screened for novel chemistry by Professor Marcel Jaspars.

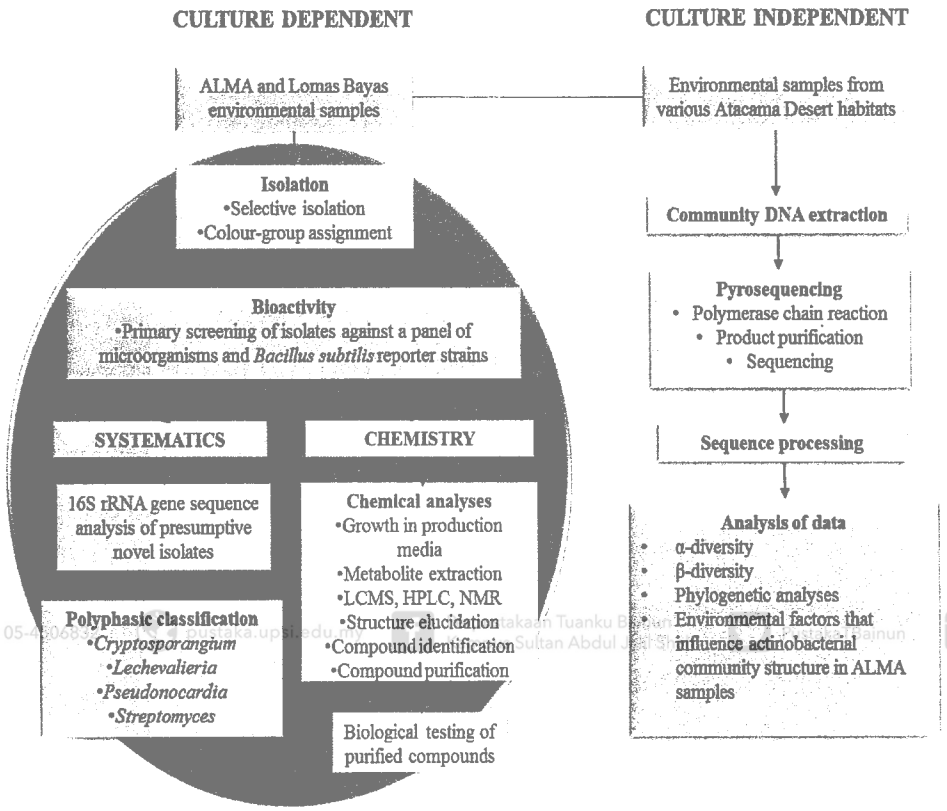

Figure 1.2 The biphasic strategy used to determine actinobacterial diversity in Atacama Desert environmental samples.

Outline of key contents included in each chapters of this thesis:

1. General Introduction
    - Background
    - Objectives and content of thesis
    - Biodiversity, bioprospecting and biogeography
    - Unexplored actinobacterial diversity as a taxonomic road map to drug discovery
    - Novel actinobacteria from desert ecosystems as a source of new drug leads
    - Culture-independent approaches to determine the extent of actinobacterial diversity in natural habitats

2. Materials and Methods
    - Sampling locations
    - Selective isolation of actinobacteria from ALMA and extreme hyper-arid Lomas Bayas environmental samples
    - Selection, maintenance and colour-group assignment
    - Comparative 16S rRNA gene sequencing studies
    - Classification of presumptively novel actinobacteria
    - Screening for bioactivity
    - Culture-independent study of actinobacterial diversity in Atacama Desert environmental samples

3. Biosystematic studies on actinobacterial strains isolated from the Cerro Chajnantor and Lomas Bayas environmental samples
    - Selective isolation of actinobacteria from the Cerro Chajnantor and Lomas Bayas environmental samples
    - Dereplication of presumptive actinobacteria
    - Screening of isolates for bioactivity
    - Classification of representative strains isolated from the Cerro Chajnantor and Lomas Bayas environmental samples based on 16S rRNA gene sequences

4. Polyphasic taxonomic studies on putatively novel actinobacteria isolated from Atacama Desert habitats
    - Selection of isolates and type strains

- Classification of *Actinomadura*, *Amycolatopsis*, *Cryptosporangium* and *Pseudonocardia* isolates
- Classification of putatively novel *Streptomyces* isolates

5. New specialized metabolites synthesized by novel actinobacteria isolated from soils of the Chajnantor Plateau
   - Screening of initial metabolite extracts
   - Novel specialized metabolites extracted from isolate H45
   - Biological testing of novel specialized metabolites
   - Polyphasic classification of isolates H9 and H45

6. Actinobacterial rare biospheres and dark matter revealed in habitats of the Chilean Atacama Desert
   - Actinobacterial taxon richness and diversity coverage
   - Taxonomic diversity
   - Habitat specificity and co-occurrence of taxa

7. Physico-chemical properties of Cerro Chajnantor environmental samples
   - Physico-chemical properties of ALMA environmental samples
   - Actinobacterial community composition
   - Actinobacterial community variation between ALMA environmental samples

8. Major outcomes and perspectives for future work

1.3 Biodiversity, bioprospecting and biogeography

Biological diversity or biodiversity includes *genetic diversity*, the variation of genes and genomes within a species (intraspecific diversity); *species diversity*, the number of species within a community, and *ecological diversity*, the number of communities in an ecosystem (Bull 2004; Harper and Hawksworth 1995). For most practical purposes the basic unit used in biodiversity studies is the species (Heywood and Baste 1995; Duelli and Obrist 2003). Microbial diversity deals with studies of archaea, bacteria, fungi, microalgae and protozoa. The simplest measure of diversity is species or α-diversity, that is the number of species within a community or habitat, in turn, β-diversity is a measure of between area diversity and γ-diversity is the extent of within-area diversity at biome

and bioregional levels (Whittaker 1972). Other ecological measures include *species abundance*, the number of individuals per species (McGill et al. 2007) and *functional diversity*, the activity of organisms in communities and ecosystems (Petchey and Gaston 2006).

Biodiversity underpins biotechnology as new products and processes depend upon the discovery, evolution and exploitation of organisms (Bull et al. 1992; Bull et al. 2000; Bull et al. 2016). The number of published prokaryotic species, which currently stands at around 13,000, is very small compared to the number of animal and plant species, somewhat less in comparison to the number of fungi and protozoa (Mora et al. 2011). Establishing the total number of prokaryotic species remains subjective and is dependent upon the matrix used to define species (Yarza et al. 2014). Yarza and his colleagues estimated the total number of prokaryotic species to be around 4×10^5, a value much lower than an earlier estimate of 2×10^6 (Curtis et al. 2002) which was just for those present in oceanic waters. These values seem low as there is evidence that all microbiomes seem to harbour significant numbers of unique prokaryotic taxa, a recent example for the human microbiome identified more than 60 prospective new species (Browne et al. 2016). It seems plausible, therefore, that each of the >1 million species of animals will support at least one novel prokaryotic species, not to mention those associated with the quarter of a million plant species.

It is now common knowledge that the vast majority of prokaryotes in nature are either uncultivable or remain to be cultivated (Staley and Konopka 1985; Torsvik et al. 2002; Ward 1998; Amann et al. 1995; Rappé and Giovannoni 2003; Oren 2004) hence an astonishing genetic and metabolic diversity is available for bioprospecting, a term introduced to cover the systematic search in nature for organisms that produce novel bioactive molecules. Such studies are increasingly focused on the isolation and screening of extremophilic and extremotolerant microorganisms isolated from biomes under extreme environmental pressures (Bull 2011; Bull et al. 2016).

Recent culture-independent data derived from 16S rRNA gene sequencing surveys show that prokaryotic diversity can be divided into two categories (de Pascale et al. 2012). On one hand, there are abundant phylotypes which form a major fraction of the total 16S rRNA gene pool, but a tiny fraction of all of the phylotypes. In contrast, there is an overwhelming number of low abundance phylotypes, the "rare biosphere" (Sogin et al. 2006; Bent and Forney 2008; Pedrós-Alió 2012; Lynch and Neufeld 2015) or "dark matter" (Rinke et al. 2013; Hedlund et al. 2014), which, despite their diversity, constitute a minor fraction of the total ribosomal gene pool. The species richness of "rare

biospheres" has still to be unravelled (Kirchman et al. 2010; Will et al. 2010; Bartram et al. 2011; Yang et al. 2012), as is its significance for bioprospecting campaigns.

Although only a tiny fraction of actinobacterial diversity from natural habitats has been cultivated and screened for bioactivity it is clear that filamentous actinobacteria are an invaluable source of medically useful antibiotics, as exemplified by the synthesis of erythromycin from *Saccharopolyspora erythraea* NRRL 4338^T (Oliynyk et al. 2007); gentamicin from "*Micromonospora purpurea*" NRRL 2953 (Weinstein et al. 1963; Wagman 1980), which is now a synonym of *Micmonospora echinospora* NBRC 13149^T (Kasai et al. 2000); teichoplanin from *Actinoplanes teichomyceticus* AB 8327^T (Somma et al. 1984)and vancomycin from *Amycolatopsis orientalis* NRRL 2430^T (Lechevalier et al. 1986; Pittenger and Brigham 1956)

Filamentous actinobacteria, especially streptomycetes, are also unusual as they are able to synthesize many bioactive metabolites, as exemplified in Table 1.1. Since chemical diversity often follows biological diversity it can be anticipated that screening novel actinobacteria for bioactivity will lead to a continuing source of new antibiotics, especially since whole-genome sequences of representatives of diverse taxa, such as the genera *Amycolatopsis, Salinispora, Saccharopolyspora, Saccharothrix* and *Streptomyces* have shown that they are rich in biosynthetic gene clusters that code for the production of unknown compounds that are expected to be bioactive (Bentley et al. 2002; Oliynyk et al. 2007; Udwary et al. 2011) (Strobel et al. 2012; Tang et al. 2015). Consequently, it can be expected that the taxonomic approach to drug discovery (Figure 1.1) will lead to the detection of new antibiotic drug leads. In addition, understanding how such organisms adapt to harsh environmental conditions, as shown by analyses of the genome of *Modestobacter caceserii* KNN 45-2b^T, a strain isolated from an extreme hyper-arid Atacama Desert soil, is an important challenge (Busarakam et al. 2016a).

Biogeography, the study of the global distribution of species across space and time, is important for understanding biodiversity, notably for establishing whether prokaryotic species are found in more than one geographical area or are restricted to a defined area, in other words whether they are cosmopolitan or endemic species (Staley and Gosink 1999; Ramette and Tiedje 2007). This subject clearly has implications for the taxonomic approach to the discovery of natural products from actinobacteria. The obligate marine genus *Salinispora* has proved to be an interesting model by which to study bacterial biogeography (Jensen et al. 2007; Freel et al. 2012; Jensen and Mafnas 2006). In wide ranging genetical and phylogenetic studies evidence was found for a cosmopolitan distribution of *Salinispora arenicola* and regional endemism for the two

remaining species, *Salinispora pacifica* and *Salinispora tropica*. The co-occurance of *S. arenicola* with the other two *Salinispora* species was seen to be evidence of ecological differentiation while comparatively high levels of intraspecific diversity were apparent in *S. pacifica*. Antony-Babu et al. (2008) provided genetic and phenotypic evidence for *Streptomyces griseus* ecovars across a beach and dune sand system thereby showing that endemism can occur within an actinobacterial species where spore dispersal is not constrained. Coupling between taxonomic and chemical diversity may occur at the level of ecovars as well as at higher taxonomic ranks (Ward and Goodfellow 2004).

Table 1.1 Approximate number of bioactive microbial metabolites in periods from 1940 to 2010 according to their producers.

Periods	*1940-1974 Early years*	*%*	*1975-2000 Mid-era*	*%*	*2001-2010 New age*	*%*	*Total*
Species							
Actinobacteria	3400	62	7200	42	3100	28.5	13700
Streptomyces ssp.	2900		5100		2400		10400
Other actinobacteria	500		2100		700		3300
All microscopic bacteria	800	15	2300	13	1100	10	4200
Myxobacteriales	25		400		210		635
Cyanobacteria	10		30		1250!		1290!
All Fungi	1300	23	7700	45	6600	61	15600
Microscopic fungi	950		5400		4900		11250
Basidiomycetes	300		1800		1500		3600
Other fungi	20		200		160		380!
Total per year	5500/180		17000/690		10800/1100		33500

Taken from Bérdy (2012).

1.4 Unexplored actinobacterial diversity as a taxonomic roadmap to drug discovery

1.4.1 Habitat selection

The first step in taxonomic approaches to bioprospecting involves the selection of the ecosystems to be sampled. Innumerable actinobacteria have been isolated and screened for bioactivity since the momentous discovery that *Streptomyces griseus* produced streptomycin (Schatz et al. 1944), the first antibiotic to be used for anti-tuberculosis therapy. Early bioprospecting campaigns built upon this discovery as they were focused mainly on the isolation and screening of *Streptomyces* strains from a range of soil types. Initially, this pragmatic approach to bioprospecting led to the discovery of many new antibiotics, including ones of therapeutic value, such as chloramphenicol from *Streptomyces venezuelae* ATCC 10712^T (Ehrlich et al. 1948) and neomycin from *Streptomyces fradiae* ATCC 10745^T (Waksman and Lechevalier 1949). This approach fell out of favour when common soil streptomycetes were repeatedly found to produce well known antibiotics.

Williams and his colleagues were amongst the first to realize that the presence, distribution, numbers and kinds of actinobacteria in natural habitats were influenced by environmental factors such as aeration, pH, temperature and the availability of organic matter and water (Williams and Mayfield 1971; Williams et al. 1972; Goodfellow and Williams 1983). It subsequently become clear that the prospect of isolating novel actinobacteria was a product of the biological properties of environmental samples, as exemplified by the presence of large numbers of novel or putatively novel acidophilic and acidotolerant actinobacteria in acid forest soils (Khan and Williams 1975; Kim et al. 2003; Golinska et al. 2015b; Golinska et al. 2013b; Golinska et al. 2013a) and their alkaliphilic counterparts in a beach and dune sand system (Antony-Babu and Goodfellow 2008; Antony-Babu et al. 2008).

The realization that novel actinobacteria were a feature of previously overlooked natural habitats promoted a shift towards the isolation and screening of novel actinobacteria from many different sources (Tiwari and Gupta 2013), notably extreme habitats (Ward and Goodfellow 2004; Bull 2011; Goodfellow et al. 2013). In practice, this strategy led to comprehensive studies on the biosynthetic potential of novel actinobacterial strains isolated from previously unexplored sources, including insects (Carr et al. 2012; Madden et al. 2013; Guo et al. 2012; Kim et al. 2014a; Bai et al. 2016;

Beemelmanns et al. 2016), roots and leaves of medicinal and tropical plants (Duangmal et al. 2008; Qin et al. 2009; Janso and Carter 2010; Golinska et al. 2015a; Wardecki et al. 2015) and, in particular, from diverse marine habitats (Bull et al. 2005; Jensen 2010; Zotchev et al. 2012; Manivasagan et al. 2014). An interesting feature of such studies is that rare actinobacteria representing taxonomically diverse genera were seen to be a rich source of new antibiotics though streptomycetes still remain supreme in this respect (Tiwari and Gupta 2012).

1.4.2 Selective isolation of actinobacteria from natural habitats

Most actinobacteria in natural habitats are saprophytes which tend to be overgrown by fungi and other bacteria on standard nutrient media hence the need to use selective procedures to isolate them from environmental samples in dilution plate experiments. It is now established practice to add antifungal antibiotics, such as cycloheximide, nystatin and pimaricin, to actinobacterial isolation media to control or eliminate the growth of fungal colonies (Porter 1960; Gregory and Lacey 1962; Williams and Davies 1965). Similarly, penicillin G and polymixin B select actinobacteria from fast-growing bacteria (Williams and Davies 1965; Nonomura and Ohara 1969), as do nalidixic acid and trimethoprim (Hayakawa et al. 1996).

In order to isolate representative strains from actinobacterial communities present in natural habitats it is necessary to use several taxon-specific isolation procedures as individual populations that constitute communities have different biological needs, including growth and incubation requirements. Specific selective isolation procedures used to isolate actinobacteria from environmental samples have been the subject of comprehensive reviews (Nolan and Cross 1988; Labeda and Shearer 1990; Goodfellow and Fiedler 2010; Tiwari and Gupta 2012). The choice of selective media for bioprospecting campaigns is somewhat subjective but is influenced by the biome to be studied, as exemplified by the need to use acidified selective media for the isolation of acidophilic actinobacteria (Khan and Williams 1975; Busti et al. 2006; Golinska et al. 2013b; Golinska et al. 2013a; Golinska et al. 2015b). Nevertheless, the selection of both macro- and micro-habitats as a source of novel metabolically active actinobacteria remains a matter of experience and judgement.

The procedures used to selectively isolate actinobacteria from environmental samples while many and varied tend to follow a number of common steps, namely

pretreatment of samples, detachment of propagules (hyphal fragments and spores) from particulate matter, inoculation and incubation of selective isolation plates and selection of representative colonies or choice of target strains for preservation and further study. The selective experimental procedures highlighted in the following sections are ones considered to be most relevant to the isolation of taxonomically diverse actinobacteria from arid desert soils, additional methods can be found in the review articles cited above.

1.4.3 Pretreatment of environmental samples

Selection of actinobacteria can be enhanced by chemical and physical pretreatment of either environmental samples or propagules in suspensions before inoculating onto selective media (Cross 1982; Goodfellow and Fiedler 2010). It is particularly important to detach actinobacterial propagules from particulate substrates and to dispense soil aggregates as microorganisms may be bound within them. The resultant suspensions are serially diluted, plated onto selective media and incubated (Williams et al. 1984). However, the physico-chemical interactions of microbial propagules with organic matter and soil particles affect their recovery from environmental samples.

Several cell dispersion procedures have been used to detach microbial propagules from particulate material (Hopkins et al. 1991; Hayakawa et al. 2000), including the use of buffered diluents (Niepold et al. 1979) and ultrasonication (Ramsay 1984). Procedures such as these address the problem of representative sampling but involve a trade-off between the extent of propagule release and cell death. The dispersion and differential centrifugation (DDC) method, a multistage procedure introduced by Hopkins et al. (1991), combines several physico-chemical treatments and has been shown to be especially effective in increasing the yield of actinobacterial propagules from samples taken from natural habitats (Hopkins et al. 1991; Macnaughton and O'Donnell 1994; Atalan et al. 2000; Sembiring et al. 2000; Maldonado et al. 2005). There is also some evidence that different *Streptomyces* species may be preferentially isolated at different stages of the DDC procedure (Atalan et al. 2000).

The effectiveness of physical pretreatment techniques depends to some extent on the biological properties of the target bacteria. In general, actinobacterial spores tend to be more resistant to desiccation than vegetative cells of bacteria hence simply air-drying soil samples greatly increases the prospect of isolating spore-forming strains (Williams et al. 1972; Hayakawa et al. 1991b; Hayakawa et al. 1991a; Whitham et al. 1993).

Resistance to desiccation is usually associated to some degree with resistance to heat, drying soil samples, for instance, held at or above 100°C for an hour reduces the number of unwanted bacteria thereby facilitating the recovery of actinobacteria (Nonomura and Ohara 1969). Actinobacterial propagules are more sensitive to dry than to wet heat which means that relatively low heat treatment regimes can be used to treat soil suspensions (Rowbotham and Cross 1977; Manfio et al. 2003). It is, however, important to recognize that while pretreatment procedures improve the ratio of actinobacterial to bacterial counts on isolation plates, the number of actinobacteria may be decreased (Williams et al. 1972).

1.4.4 Choice of selective isolation media

Selective media are designed to support the growth of target microorganisms at the expense of unwanted ones. Media selectivity is an expression of nutrient competition, pH, presence of inhibiting compounds and incubation conditions. The choice of selective isolation media can be critical, especially in bioprospecting campaigns designed to isolate representative samples of actinobacterial communities from natural habitats. Many 'non-selective' media have been formulated for the isolation of broad ranges of actinobacterial taxa (Williams et al. 1984; Goodfellow and Fiedler 2010). Surprisingly, many such media were designed without reference to the biological properties of actinobacteria *per se*, as illustrated by the use of colloidal chitin (Lingappa and Lockwood 1962; Hsu and Lockwood 1975) and starch-casein agars (Küster and Williams 1964). These media are now known to select for a relatively narrow range of *Streptomyces* species (Vickers et al. 1984; Williams et al. 1984), but are still widely used to detect the presence of streptomycetes in extreme habitats where these organisms may be the major component of prokaryotic communities (Pathom-Aree et al. 2006a; Okoro et al. 2009). In contrast, arginine-vitamin (Nonomura and Ohara 1969) and humic acid agars (Hayakawa and Nonomura 1987) support the recovery of taxonomically diverse actinobacteria from poorly studied biomes (Whitham et al. 1993; Busarakam 2014; Wang et al. 2015; Busarakam et al. 2016a).

The most effective taxon-specific selective isolation strategies are based on the biological properties of the target actinobacteria, notably using information on nutrient and antibiotic sensitivity profiles drawn from phenotypic databases (Goodfellow and Haynes 1984; Goodfellow and O'Donnell 1994; Goodfellow and Fiedler 2010; Williams and Vickers 1988). Selective media have been recommended for the isolation of many

actinobacterial genera (Goodfellow and Fiedler 2010; Tiwari and Gupta 2012), including ones found to support the isolation of novel or rare species of *Actinomadura* (Athalye et al. 1981; Busarakam 2014), *Amycolatopsis* (Tan 2002; Busarakam et al. 2016a), *Dactylosporangium* (Kim et al. 2011), *Nocardia* (Orchard and Goodfellow 1974, 1980) and *Streptomyces* (Vickers et al. 1984; Busarakam 2014). The effectiveness of such selective isolation media is often enhanced by the use of appropriate pretreatment regimes (Labeda and Shearer 1990; Goodfellow and Fiedler 2010).

It is obvious that incubation conditions will contribute to selectivity as the growth of actinobacteria is influenced by gaseous and temperature requirements and by the length of incubation. Inoculation of selective isolation plates at 4°C, 25°C and 45°C favours the isolation of psychrophilic, mesophilic and thermophilic actinobacteria, respectively. Incubation for up to five weeks may be required to isolate members of the families *Micromonosporaceae* and *Streptosporangiceae* whereas *Nocardiaceae* and *Streptomycetaceae* strains generally grow well on selective isolation media after 7 to 10 days (Goodfellow and Fiedler 2010; Labeda and Shearer 1990).

1.4.5 Colony selection

The selection of actinobacterial colonies growing on selective isolation plates is one of the most time-consuming and subjective stages of the culture-dependent bioprospecting strategy. Colonies can be selected randomly if a broad-range of actinobacteria are sought or with a degree of objectivity where target-specific isolation media are used. Actinobacteria can be assigned to target taxa on the basis of characteristic colonial features and by examining colonies directly on selective isolation media using a long working distance objective to detect diagnostic morphological features, such as the presence of spore vesicles or spore chain morphology. Streptomycetes, for instance, produce characteristic filamentous leathery colonies covered by an abundant aerial spore mass which may be pigmented whereas *Amycolatopsis* strains form branched filamentous colonies covered by powdery white sterile aerial hyphae (Atalan et al. 2000; Tan et al. 2006b; Busarakam et al. 2016a). However, it is rarely possible to distinguish between species of actinobacterial genera growing on selective isolation plates. In such instances a sample of the various colony types need to be subcultured and preserved for further study. It is also good practice to select unusual colonies for further study as they may prove to be members of novel taxa.

1.4.6 Dereplication of isolates

Dereplication, a critical stage in the culture-dependent bioprospecting pipeline, is designed to rapidly assign actinobacteria with similar phenotypic properties to taxonomically meaningful groups so that representative isolates can be selected for screening assays. An increasing number of dereplication procedures are available for this purpose, including chemotaxonomic and molecular fingerprinting techniques (Schumann and Maier 2014; Maldonado et al. 2008; Ferguson et al. 1997). An effective way of dereplicating *Streptomyces* strains was introduced by (Williams et al. 1969) who assigned large numbers of these organisms to groups based on aerial spore mass, substrate mycelial and diffusible pigment colours formed on oatmeal agar and on their ability to produce melanin pigments on yeast extract- malt extract- iron agar. Colour-groups were later shown to be a measure of streptomycete diversity in rhizosphere and non-rhizosphere soils (Williams and Vickers 1988; Atalan et al. 2000; Sembiring et al. 2000), as colour-group representatives were found to belong either to validly published or novel *Streptomyces* species based on computer-assisted identification (Williams and Vickers 1988; Atalan et al. 2000) and polyphasic taxonomic procedures (Manfio et al. 2003; Goodfellow et al. 2007). It is encouraging that a linear correlation has been found between streptomycete colour-groups and corresponding *rep*-PCR data (Antony-Babu et al. 2010).

The assignment of streptomycetes to colour-groups has provided an insight into their taxonomic diversity in a beach and dune sand ecosystem (Antony-Babu and Goodfellow 2008), in desert soils (Okoro et al. 2009; Busarakam 2014) and marine sediments (Goodfellow and Haynes 1984; Pathom-Aree et al. 2006a) thereby facilitating the choice of representative isolates for bioactivity assays (Goodfellow and Fiedler 2010; Busarakam 2014) . It is particularly significant that this approach has led to a marked increase in hit rates in screening assays (Goodfellow and Fiedler 2010) .

1.4.7 Screening for bioactivity

Primary screening of dereplicated actinobacteria can be achieved using standard "kill" assays in order to detect compounds active against panels of Gram-positive and Gram-negative bacteria and in some cases yeasts (Fiedler 2004; Baltz 2007). In practice, this often simply involves taking agar plugs from actinobacterial lawns and placing them onto

corresponding lawns of wild type strains, incubating overnight then recording zones of inhibition. Variants of this approach include spotting culture filtrates or organic extracts directly onto wild type strains and examining for inhibition of growth either by eye or spectrophometerically (2008; Bredholdt et al. 2007; Hong et al. 2009). Hits can be prioritized according to the taxonomic novelty of the actinobacterial strains using rDNA sequence comparisons to known actinobacteria. Further prioritization can be achieved by detecting the modes of action (MOA) of bioactive agents in secondary screens (Fiedler 2004).

An array of genetic constructs are available to detect the MOA of different classes of antibiotics (2004b); Hutter et al. (2004a). Mutant strains can be designed to carry reporters, such as β-galactosidase and luciferase genes, which are fixed to promoters that respond to specific antibiotics that trigger the expression of the reporter gene (Hutter et al. 2004b; Urban et al. 2007). The production of β-galactosidase, for instance, leads to the colourless 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal) being cleaved into β-D-galactose and the blue-coloured 5-bromo-4-chloro-3-hydroxyindole (Figure 1.3). The formation of the blue pigment shows that the bioactive compound is repressing the target pathway. *Bacillus subtilis* reporter strains have been used to detect inhibitors of cell wall, DNA, fatty acid and protein biosynthesis (Hutter et al. 2004a; Busarakam 2014).

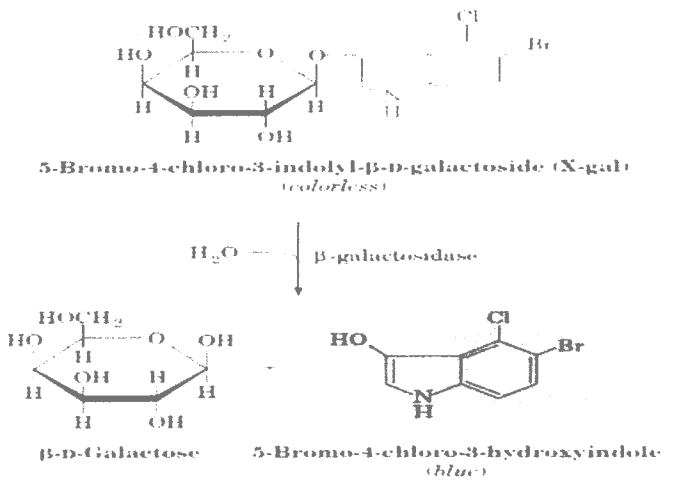

Figure 1.3 Cleavage of 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal) to β-D-galactose and 5-bromo-4-chloro-3-hydroxyindole (Blue compound). (Adapted from (Voet and Voet 2011).

1.4.8 Detection and classification of novel actinobacterial species

The quality of actinobacterial isolates examined in bioprospecting campaigns is a function of two parameters, diversity and dereplication, each of which is much more important than the overall number of strains. Diversity in this context is measured in its taxonomic sense, namely how many different species, genera and families make up the strain selection. The more taxonomically diverse the collection the greater the chance of finding novel specialized metabolites. Dereplication, as stated earlier, is the assignment of isolates to taxonomically predictive groups. Bioprospecting studies still tend to be based on large numbers of strains which include multiple repeats of clonal or closely related isolates which makes for inefficient screens.

The taxonomic approach to culture-dependent bioprospecting may take one of several forms depending on the nature of the exercise. For instance, dereplicated isolates shown to belong to novel taxa can be screened for bioactivity or representative strains selected from isolation media can be screened and those considered interesting checked for novelty. Either way diversity and dereplication can be achieved using various taxonomic measures. Presently the method of choice is 16S ribosomal RNA gene sequencing (Goodfellow and Fiedler 2010; Yarza et al. 2014; Yarza and Munoz 2014), though this can be expected to change as whole-genome sequencing becomes more prevalent (Klenk and Göker 2010; Mende et al. 2013; Sangal et al. 2014; Thompson et al. 2014). Nevertheless, it is still a requirement to describe novel species of prokaryotes using a combination of genotypic and phenotypic features (Wayne et al. 1987; Kämpfer and Glaeser 2012; Vandamme et al. 1996; Gillis et al. 2005) and to follow the rules embodied in the *International Code of Nomenclature for Bacteria* (Lapage et al. 1975; Lapage et al. 1992).

The current strategy used to classify prokaryotes is based on the integrated use of taxonomic information generated from the application of chemotaxonomic, molecular systematic and phenotypic methods (Vandamme et al. 1996; Schleifer 2009; Goodfellow 2000). This polyphasic approach has led to major improvements in the classification of archaea and bacteria, not least in establishing phylogenetic relationships within and between actinobacterial taxa (Ludwig et al. 2012; Girard et al. 2013). The choice of genotype and phenotypic methods for any particular study is critical as they need to reflect the rank and biological properties of taxa and the use of new procedures that provide improved resolution between taxa (Tindall et al. 2010; Cody et al. 2014; Vandamme and

Peeters 2014; Schumann and Maier 2014; Kim et al. 2014b). Methods found to be useful in the classification of actinobacteria are shown in Table 1.2.

The evolutionary conserved nature, ubiquitous distribution and size of 16S rRNA genes was instrumental in them becoming the phylogenetic marker of choice following the pioneering studies on the universal tree of life by Karl Woese and his colleagues (Woese and Fox 1977; Woese 1987). In essence, the use of such sequences in the classification and identification of prokaryotes rests on comparisons of new sequences against comprehensive databases of known sequences, notably those of type strains, as illustrated by Chun and his colleagues (Chun et al. 2007; Kim et al. 2012b; Kim and Chun 2014). Extensive comperative 16S rRNA gene sequencing studies have revolutionized our understanding of phylogenetic relationships among prokaryotes (Yarza and Munoz 2014; 2010; Yarza et al. 2008). Indeed, the framework for the classification of archaea and bacteria in the current edition of *Bergey's Manual of Systematic Bacteriology* is based on comparative 16S rRNA gene sequence data, as shown by the assignment of actinobacteria to 6 classes, 23 orders, 53 families and 222 genera (Goodfellow et al. 2012a) . The procedures used to generate, align and analyse high quality 16S rRNA gene sequences are not considered here as they have been covered in recent reviews (Rosselló-Móra et al. 2011; Ludwig et al. 2011).

Strain prioritization based on comparative 16S rRNA gene sequences is an important stage in the culture-dependent approach to natural product discovery at a time when actinobacterial strain culture collections are rapidly expanding in light of improved selective isolation procedures. Isolates representing dereplicated taxa, such as *Streptomyces* colour-groups, found to form distinct lineages in the actinobacterial 16S rRNA gene tree can be prioritized for screening, especially for low throughput screens. Alternatively, isolates found to produce new metabolites in more broadly-based screening programmes can be checked for novel sequences using digital databases such as the EzTaxon platform (Kim and Chun 2014). Either way putatively novel isolates should be examined further to establish whether they merit publication as new taxa. The importance of publishing new internationally recognised names for commercially significant prokaryotes is often overlooked even though an organism's name is a key to all that is known about it (Goodfellow and Fiedler 2010).

Table 1.2 Sources of taxonomic information for the classification of actinobacteria.

Cell component		Taxonomic rank		
		Genus or above	Species	Subspecies or below
Chemical markers	Fatty acids	/	/	
	Isoprenoid quinones	/	/	
	Mycolic acids	/	/	
	Peptidoglycans	/		
	Polar lipids	/	/	
	Polysaccharides	/	/	
Chromosomal DNA	Base composition (mol % G+C)	/	/	
	DNA:DNA hybridisation		/	/
	Restriction patterns (AFLP, PFGE, RFLP, ribotyping)		/	/
	Whole-genome analyses	/	/	/
DNA segments	DNA probes	/	/	/
	16S rRNA gene sequencing	/	/	/
	Multi-locus sequence analyses		/	/
	PCR-based DNA fingerprinting (PCR-RFLP, ITS, RAPD, Repetitive-PCR		/	/
Expressed features	Metabolism		/	
	Morphology	/	/	
	Physiology	/	/	
Proteins	Amino acid sequence analyses	/	/	
	Multilocus sequence typing	/	/	/
Ribosomal RNA	Restriction patterns (ARDRA)		/	/
	t-RNA fingerprinting		/	/
Whole organism	MALDI-TOF mass spectrometry		/	/

Abbreviations: AFLP, amplified fragment length polymorphism; ARDRA, amplified ribosomal DNA restriction analyses; ITS, intergenic 16S-23S rRNA transcribed spacer sequences; PFGE, pulse field gel electrophoresis; RFLP, restriction fragment length polymorphism; RAPD, randomly amplified polymorphic DNA fingerprinting; PCR, polymerase chain reaction.

In general, the discontinuous distribution of chemotaxonomic and morphological markers underpin actinobacterial phylogenies based on 16S rRNA gene sequence data (Goodfellow 2012a; Ludwig et al. 2012). *Streptomyces* strains, for example, form an extensively branched substrate mycelium, aerial hyphae that typically differentiate into chains of spores, have LL-diaminopimelic acid (LL-A_2pm) but no diagnostic sugars in the peptidoglycan, contain major proportions of saturated, *iso*- and branched chain fatty acids, usually have either hexa- or octahydrogenated menaquinones (MK-9[H_6], MK9[H_8]) as the predominant isoprenologue and complex polar lipid patterns rich in diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol and phosphatidylinositol mannosides (Kämpfer 2012). In contrast, *Pseudonocardia* strains form substrate and aerial mycelia with spore chains produced by acropetal budding or fragmentation, have *meso*-A_2pm, arabinose and galactose in the peptidoglycan, tetrahydrogenated menaquinones with eight isoprene units (MK8[H_4]) as the major isoprenologue, *iso*-branched hexadecanoic acid as the predominant fatty acid and either phosphatidylcholine or phosphatidylethanolamine as diagnostic polar lipids (Huang and Goodfellow 2015).

In short, discontinuously distributed markers such as those outlined above provide an effective means of evaluating actinobacterial phylogenies, especially in instancies where phylogenetic relationships do not allow groups to be recognised with confidence (Labeda et al. 2011) or where the use of different treeing algorithms yield different phylogenies (Ludwig and Klenk 2005; Ludwig et al. 2012; Kämpfer and Glaeser 2012). Standard procedures are widely used to detect chemical markers found to be of value in actinobacterial systematics (Table 1.3), notably protocols for the detection of diaminopimelic acid isomers (Staneck and Roberts 1974; Hasegawa et al. 1983), fatty acids (Kroppenstedt 1985), including mycolic acids (Minnikin et al. 1980), and menaquinones and polar lipids (Minnikin et al. 1984).

Several molecular systematic methods are available to assign closely related bacterial strains to species where insufficient variation is present in 16S rRNA gene sequences (Wayne et al. 1987; Stackebrandt et al. 2002). The genus *Streptomyces* is a good case in point as many multimembered lineages recognised in an extensive comparison of 16S rRNA gene sequences of type strains were not supported by high bootstrap values (Labeda 2011). In such instances, DNA:DNA hydridisation (DDH) studies allow closely related strains to be clustered into genomic species, that is, species that encompass strains that have approximately 70% or more DNA relatedness with <5°C stability in pairwise heteroduplices (Wayne et al. 1987). DDH assays are extensively used

to delineate actinobacterial species, including *Streptomyces* species (Labeda and Lyons 1991a, b; Labeda 1998; Rong and Huang 2012) even though they are tedious to carry out and subject to experimental error (Gevers et al. 2005; Rosselló-Móra et al. 2011), the reliability of the different DNA:DNA reassociation methods has been evaluated by (Rosselló-Mora 2006). It is now good practise to carry out DDH experiments only when strains share high 16S rRNA gene sequence similarities, recommended thresholds have risen from 97.0% (Stackebrandt and Goebel 1994) to 99.0% (Meier-Kolthoff et al. 2013) then back to 98.7% (Kim et al. 2014b) though cut off values vary depending on the taxa under study (Stackebrandt and Ebers 2006; Meier-Kolthoff et al. 2013).

Stackebrandt et al. (2002) recommended the application of DNA fingerprinting and multilocus sequence analyses (MLSA) to distinguish between closely related bacterial species. Comparative MLSA is a product of multilocus sequence typing that has been widely used in epidemiological surveys of medically significant taxa (Cody et al. 2014) and shown to be of value in unravelling species diversty between actinobacteria that share very high 16S rRNA gene sequences (Ludwig et al. 2012). The underlying rationale and steps involved in MLSA analyses have been described in detail (Cody et al. 2014; Rong and Huang 2012). In essence, MLSA is based on the generation and analyses of 400-500 base pair sequences of fragments of five to seven housekeeping genes and the presentation of the resultant data in concatenated gene trees. Such trees show greater resolution between closely related actinobacteria than corresponding 16S rRNA gene trees, as evidenced by comparative studies on the genera *Kribbella* (Curtis and Meyers 2012), *Micromonospora* (Carro et al. 2012) and *Salinispora* (Freel et al. 2012), and extensive investigations on phytopathogenic (Labeda et al. 2014; Labeda 2011) and commercially significant (Guo *et al.,* 2008; Rong & Huang, 2010, 2012; Rong *et al.,* 2010, 2012) streptomycetes. A strong correlation between MLSA, DDH and phenotype data was highlighted by (Rong and Huang 2014) who also noted that the DDH 70% threshold corresponded to a five gene sequence of 0.007. MLSA data can also be used to resolve the taxonomic status of isolates that form distinct branches in unstable regions of the 16S rRNA ene tree, as exemplified by the classification of the type strain of *Streptomyces leeuwenhoekii* (Busarakam et al. 2014).

The advent of rapid and relatively inexpensive whole-organism sequencing (WGS) technologies and associated bioinformatic tools are leading to a step change in the way prokaryotic systematics is being conducted (Klenk and Göker 2010; Chun and Rainey 2014; Harris and Okoro 2014), not least the introduction of new parameters for establishing species boundaries (Table 1.3). The average nucleotide identity (ANI) is a

particularly attractive metric for the delineation of species with a 94% ANI value corresponding to the 70% DDH cut-off (Richter and Rosselló-Móra 2009; Arahal 2014), WGS data also provide useful insights into an organism's ecological, metabolic and biotechnological potential (Wu et al. 2012; Doroghazi and Metcalf 2013; Sangal et al. 2014) and into phenomena such as gene duplication and horizontal gene transfer (Kamneva and Ward 2014). However, it is important that taxonomies based on WGS data follow sound taxonomic practices, such as the use of the nomenclatural type concept and the requirement to deposit type strains in two service culture collections in different countries (Lapage et al. 1992; Stackebrandt et al. 2014). A case has also been made for the continued use of phenotypic data in prokaryotic systematics (Kämpfer et al. 2014). In contrast, strong arguments have been presented to revise the principles and practices of prokaryotic systematics in light of the new development, not least the urgent need to streamline procedures for the publication of new taxa (Sutcliffe *et al.,* 2012 (Sutcliffe et al. 2012; Vandamme and Peeters 2014; Thompson et al. 2014; Whitman 2014).

Comparative genome-based analyses have been used to unravel actinobacterial relationships that proved difficult to resolve using established taxonomic procedures. It has, for instance, been shown that the genus *Kitasatospora* merits recognition as a sister genus to *Streptomyces* in the family *Streptomycetaceae (Girard et al. 2014; Girard et al. 2013)* and that the genera *Amycolatopsis* and *Rhodococcus* encompass multiple lineages thereby providing a framework for their taxonomic revision (Creason et al. 2014; Sangal et al. 2014; Tang et al. 2015). Whole- genome sequences are rich in genes that provide an understanding of the ecophysiology of individual strains, as in the case of *Micromonospora lupini* Lupac 08 and *Modestobacter caceresii* KNN44-2b^T, strains isolated from nitrogen-fixing nodules of *Lupidus angustifolious* and extreme hyper-arid Atacama Desert soil, respectively (Trujillo et al. 2007; Busarakam et al. 2016b). Similarly, comparative genomics has revealed evidence of marine adaptation in representatives of *Salinispora* species (Penn and Jensen 2012). An innovative approach to natural product discovery based on diverse actinobacterial WGS data showed a correlation between phylogeny and the discontinuous distribution of families of biosynthetic gene clusters (Doroghazi et al. 2014), an observation that opens up the prospect of a focused taxonomic approach to genome mining for natural products (Challis 2014).

Table 1.3 Bioinformatic tools and resources used to calculate genome-based tresholds for the circumscription of prokaryotic species.

Genome relatedness index	Threshold for delineation of species	Gene finding/annotation required?	Bioinformatic tools	Description	URL
ANI	95-96%	No	Jspecies	Java-based standardize software that calculates ANIb and ANIm (Richter & Roselló-Móra, 2009)	http://imedea.uib-csic.es/jspecies/
			EzGenome	Web-based resource for ANIb calculation	http://www.ezbiocloud.net/ezgenome
GBDP	0.258	No	Genome to Genome Database Calculator	Web-based resource for calculation of pairwise, GDP distance using NCBI-BLAST, BLAT, BLASTZ and MUMner (Meier-Kolthoff *et al.,* 2013)	http://www.gqdc.dsmz.de
MUMi	None available	No	MUMi	Web-based resource for calculation of MUMi (Deloger *et al.,* 2009)	http://genome.jouy.inra.fr/mumi/
Nucleotide identity	96.5%	Yes	specI	Web-based resource and Linux standardize program for identification using 40 universal, single-copy marker genes (Mende *et al.,* 2013)	http://vm-lux.embl.de/~mende/specI//

Abbreviations: ANI, average nucleotide identity (see Arahal, 2014); GBDP, genome BLAST distance phylogeny (Henz et al. 2005); MUMi, another distance type index (see Deloger *et al.,* 2009). Modified from Chun and Rainey (2014).

Several chemical procedures can be used to detect patterns of specialized metabolites in extracts of dereplicated strains (Goodfellow and Fiedler 2010). A critical step, dereplication of extracts can be achieved using LC-MS and chemoinformatic-based approaches. The laboratory information management system (NAPIS) is now used extensively to select the most active and novel extracts for downstream processing (Jaspars, pers. comm.). HPLC-MS analyses are generally used to confirm the novelty of key specialized metabolites. The final steps in the process, scale-up fermentation and structural elucidation of pure compounds vary depending on the nature of the strains and bioactive agents, as exemplified by the discovery of the chaxamycins (Rateb et al. 2011a) and chaxapeptin (Elsayed et al. 2015).

1.5 Novel actinobacteria from desert ecosystems as a source of new drug leads

Non-polar deserts. The likelihood of isolating novel actinobacteria that produced new specialized metabolites is enhanced when bioprospecting strategies are focused on neglected and unexplored ecosystems (Goodfellow and Fiedler 2010; Tiwari and Gupta 2012), such as those found in desert regions. Deserts cover nearly 20% of the Earth's land surface but have received little attention from microbiologists until recently (Bhatnagar and Bhatnagar 2005; Bull et al. 2016). However, it is well known that living conditions in desert habitats are challenging for microorganisms due to the lack of liquid water, low organic carbon, high UV radiation and large fluctuations in temperature. Such harsh environmental conditions can be expected to favour the evolution of novel actinobacteria with new chemistries (Okoro et al. 2009; Goodfellow et al. 2013).

Small numbers of taxonomically diverse actinobacteria have been isolated from non-polar deserts, including the Mongolian (Zenova et al. 2009; Kurapova et al. 2012), Namibia (Wink et al. 2003a), Sahara (Meklat et al. 2011; Saker et al. 2015; Saker et al. 2014) , Taklamakan (Luo et al. 2012b) and Thar (Tiwari et al. 2015; Harwani 2013) deserts, notably rare actinobacteria belonging to genera classified in the families *Geodermatophilaceae (Blastococcus, Geodermatophilus* and *Modestobacter*), *Pseudonocardiaceae (Dactylosporangium, Prauserella, Saccharopolyspora* and *Saccharothrix)*, *Streptosporangiaceae (Microbispora, Microtetraspora, Nonomuraea* and *Streptosporangium)* and *Thermomonosporaceae (Actinomadura* and *Sprillospora*). Isolates from desert soils validly published as new genera are few, but include *Geodermatophilus* (Luedemann 1968), *Jiangella* (Song et al. 2005) and *Yuhushiella* (Mao et al. 2011). In contrast, the number of desert isolates validly published as new species is rising fast, as illustrated by *Amycolatopsis deserti*

from an arid Australian composite soil (Busarakam et al. 2016a), *Mycetocola manganoxydans* and *Nocardioides deserti* from the Taklamakan Desert, in Xinjiang Province, China (Luo et al. 2012b; Tuo et al. 2015) and *Prauserella isguenensis*, *Saccharothrix algeriensis* and *Saccharothrix saharensis* from the Algerian Sahara (Zitouni et al. 2004b; Saker et al. 2015; Boubetra et al. 2013). Specialized metabolites produced by some of these strains include the new anthracycline antibiotic, mutactimycin PR (Zitouni et al. 2004a) and novel dithiolopyroline analogues (Bouras et al. 2008).

The Atacama Desert. The most extensive surveys of culturable actinobacterial diversity in desert biomes have been concentrated on sites in the Atacama Desert in Northern Chile (Okoro et al. 2009; Busarakam 2014). The location and abiotic conditions associated with this desert have been considered in several reviews (Gómez-Silva et al. 2008; Azua-Bustos et al. 2012; Bull et al. 2016; Bull and Asenjo 2013) and will not be considered here. Bull and his colleagues explained that the uniqueness of the Atacama Desert resides "*in its amazing diversity of ecological niches, its geology and geochemistry, its elevation and topography and its radiation intensity*". Until recently, the significance of such features was lost on biologists as the prevailing view held sway that the harsh conditions in much of the desert, notably extreme dryness, very low organic carbon, presence of organic oxidants and intense UV-radiation made any form of life virtually impossible (McKay et al. 2004; Navarro-González et al. 2003). Subsequently, pioneering studies carried out by Chilean microbiologists showed that microorganisms were able to colonise several Atacama Desert habitats (Gómez-Silva et al. 1990; Prado et al. 1993; Campos et al. 2009). It is now clear that different kinds of microorganisms have become adapted to the extreme conditions that prevail in diverse habitats within the desert (Azua-Bustos et al. 2012; DiRuggiero et al. 2013; Paulino-Lima et al. 2013; Ziolkowski et al. 2013).

Microbial surveys of Atacama Desert soils and regoliths have been focused mainly on the isolation of strains from hyper-arid and extreme hyper-arid regions, that is, where the mean annual rainfall to mean annual evaporation is 0.05 and 0.002% respectively (Houston 2006). Very low counts of heterotrophic bacteria have been recorded from such soils, notably from the Yungay region, one of the driest areas on Earth (Navarro-González et al. 2003). Actinobacteria were first isolated from Atacama Desert soils around fifty years ago (Cameron et al. 1966; Opfell and Zebal 1967) but despite these early leads they rarely featured in subsequent broadly-based culture-dependent surveys, exceptions include the isolation of *Arthrobacter, Kocuria, Cellulomonas* and *Geodermatophilus* strains from the Yungay region

(Gómez-Silva et al. 2008), putatively novel *Geodermatophilus* and *Streptomyces* species from one of the driest areas in the Atacama Desert at Maria Elena South 275 km northeast of the Yungay region (Azua-Bustos et al. 2015) and *Nocardia, Micromonospora, Prauserella* and *Streptomyces* strains from the Tarapaci and Antofagasta regions of the desert (Bull et al. 2016).

Recently, extensive bioprospecting campaigns employing selective isolation, screening and polyphasic taxonomic procedures have shown that low counts of taxonomically diverse actinobacteria that show bioactivity can be recovered from hyper-arid and extreme hyper-arid soil taken from the Salar de Atacama and Yungay regions of the Atacama Desert, respectively (Okoro et al. 2009; Busarakam 2014). Dereplicated isolates were mainly found to be putatively novel *Streptomyces* species though some strains were shown to form distinct phyletic lines within the evolutionary radiation occupied by rare genera, including *Actinomadura, Couchiplanes, Kribbella, Lechevalieria, Modestobacter, Nonomuraea* and *Saccharothrix*. Subsequently, some of these isolates were classified as new species of *Modestobacter* (Busarakam et al. 2016b), *Lechevalieria* (Okoro et al. 2010)and *Streptomyces* (Santhanam *et al.,* 2012a, b, 2013), notably *Streptomyces leeuwenhoekii* C34^{T} (Busarakam et al. 2014). In contrast, representatives of thermophilic actinobacteria isolated from Salar de Atacama soil were found to be *bona fide* members of two rare species, *Amycolatopsis ruanii* and *Amycolatopsis thermalba* (Busarakam et al. 2016a).

Initial studies on the natural product chemistry of taxonomically novel strains isolated from Atacama Desert soils are encouraging given the discovery of novel chemical entities with biological activity (Table 1.4). It is particularly interesting that *S. leeuwenhoekii* strains are to the fore in this respect as the high quality non-fragmented genomic sequence of the type strain of this species contains 35 clusters that apparently encode for the biosynthesis of many novel and undescribed specialized metabolites (Gomez-Escribano et al. 2015). In contrast, the whole-genome sequence of *Modestobacter caceserii* KNN45-2b^{T} contains few biosynthetic gene clusters (Busarakam et al. 2016b) hence members of this taxon are unlikely to be a source of new antibiotics though they may produce melanin compounds that may act as sunscreens. The putatively novel *Streptomyces* strain isolated from a Salar de Tara high altitude soil (4,500m) in the Chilean highlands by Schulz et al. (2011) shows inhibitory activity against type 4 phosphodieterase, a discovery that might prove useful in the treatment of inflammatory diseases such as asthma.

Table 1.4 Novel specialized metabolites produced by actinobacteria isolated from Atacama Desert soils.

Organisms	Products	Structures	Bioactivity	References
Modestobacter sp. strain KNN45-2b	Peptide	Known compounds such as anthranilic acid ethyl ester and 3-β- indolacrylic acid	None	Busarakam *et al.*, unpublished
Streptomyces leeuwenhoekii C58	Atacamycins	22-membered macrolactone polyketides	Antibacteria, antitumor, inhibits activity against phosphodiesterase types 2 and 4	(Nachtigall et al. 2011; Elsayed et al. 2015)
Streptomyces leeuwenhoekii C34^T	Chaxalactones	22-membered macrolactone polyketides	Antibacteria	(Rateb et al. 2011b)
Streptomyces leeuwenhoekii C34^T	Chaxamycins	Ansamycin polyketide	Antibacteria, anticancer	(Rateb et al. 2011a)
Streptomyces leeuwenhoekii C58	Chaxapeptin	Lasso peptide	Anticancer	(Elsayed et al. 2015)
Streptomyces leeuwenhoekii C79	Four specialized metabolites	ND	Antibacteria	Fiedler *et al.*, unpublished
Streptomyces sp. strain DB634	Ataquinones	ND	Antimicrobial, inhibits activity against phosphodiesterase type 4	(Schulz et al. 2011)

ND, not determined.

1.6 Culture-independent approach to determine the extent of actinobacterial diversity

1.6.1 Detecting the unseen majority

A strong case can be made for the application of effective procedures for the selective isolation, dereplication and classification of actinobacteria from environmental samples in order to select high quality strains for biotechnological purposes. Culture-dependent strategies have shown that actinobacterial taxa once seen to be rare in natural habitats are an integral part of actinobacterial communities, as shown by the isolation and characterization of novel *Actinospica, Amycolatopsis* and *Dactylosporangium* species (Tan et al. 2006a; Kim et al. 2011; Golinska et al. 2015b). It has also been found that representatives of novel actinobacterial species are a rich source of bioactive compounds (Goodfellow and Fiedler 2010; Jensen 2010; Becerril-Espinosa et al. 2013).

A dramatically different approach to estimating the extent of prokaryotic diversity in natural habitats arose through the application of culture-independent molecular methods, including fluorescent *in situ* hybridisation (Moter and Göbel 2000; Amann and Fuchs 2008) and denaturing gradient gel electrophoresis (DGGE) (Muyzer et al. 1993), but most notably by the recognition of previously unknown taxa through the use of oligonucleotide primers designed to amplify 16S rRNA gene sequences present in extracted community DNA (Olsen et al. 1986; Pace et al. 1986) . The strengths and weaknesses of the early culture-independent molecular-based methods will not be considered here as they are covered in detail in several comprehensive reviews (Head et al. 1998; Bull et al. 2000; O'Donnell et al. 1993). Culture-independent studies clearly show that standard culture-dependent strategies grossly underestimated the extent of prokaryotic diversity in environmental samples (Ward et al. 1990; Amann et al. 1995; Suzuki et al. 1997; Vartoukian et al. 2010). Culture-dependent procedures have also seriously underestimated the degree of actinobacterial diversity in samples taken from marine and terrestrial habitats (2003b; Stach et al. 2003a).

In more recent times high throughput sequencing platforms and associated analytical tools are being used to generate bacterial genomes (Harris and Okoro 2014; Sangal et al. 2014). The strengths and weaknesses of the 454 GS FLX+ (Roche), 550xl SOLID (Life Technologies), HiSeq 2000/2500 (Illumina) and PacBio RS (Pacific

Biosciences) platforms and associated sample preparation and template amplification technologies have been considered in detail (Claesson et al. 2010; 2012b; Loman et al. 2012a; Liu et al. 2012; Luo et al. 2012a). The size of sequencer reads and coverage are platform specific. The Illumina platform is the most economic and pair-end and mate-end reads yield, highly readable assemblies with high coverage while the PacBio platform gives much larger reads that can generate genomes without any gaps. All of the platforms can be used to generate data requited for assembling draft genome sequences. Similarly, targeted sequencing of 16S rRNA genes for taxonomic and metagenomics studies can be carried out on any of the sequencing platforms.

Roche 454 sequencing, now referred to as 454-pyrosequencing, is especially useful for sequencing short DNA fragments as it allows high coverage per sample (Claesson *et al.,* 2009) and many additional samples per run (Sogin et al. 2006; Hamady et al. 2008; Rothberg and Leamon 2008). This method has been shown to be accurate, flexible and amendable to automation and has the added advantages of not requiring the generation of clone libraries, the need for labelled primers or gel electrophoresis. The principles underpinning pyrosequencing can be found elsewhere (Ronaghi 2001; Fakruddin et al. 2013). In short, pyrosequencing is based on the luminescence detection of pyrophosphate released during DNA synthesis. A schematic outline of the procedure is shown in Figure 1.4.

High throughput 16S rRNA pyrosequencing technologies have given an insight into the relative abundance and diversity of prokaryotes in several natural habitats, including Brazilian savanna soils (Rampelotto et al. 2013), desert sand (Neilson et al. 2012; An et al. 2013; An et al. 2015), glacial surfaces (Liu et al. 2015), permafrost soils (Yang et al. 2012), surface and subsurface peat layers (Serkebaeva et al. 2013) and cold seeps in the Red Sea (Yang et al. 2015). Such studies have not only highlighted bacterial diversity, including that of many unknown taxa, but have shown that microbial community structure is influenced by environmental factors, temporal differences and land use procedures. In a more restricted study, 454 pyrosequencing revealed the relative abundance and diversity of bacterial predators, namely *Bdellovibrio* and related organisms, in fresh and salt water (Li and Williams 2015).

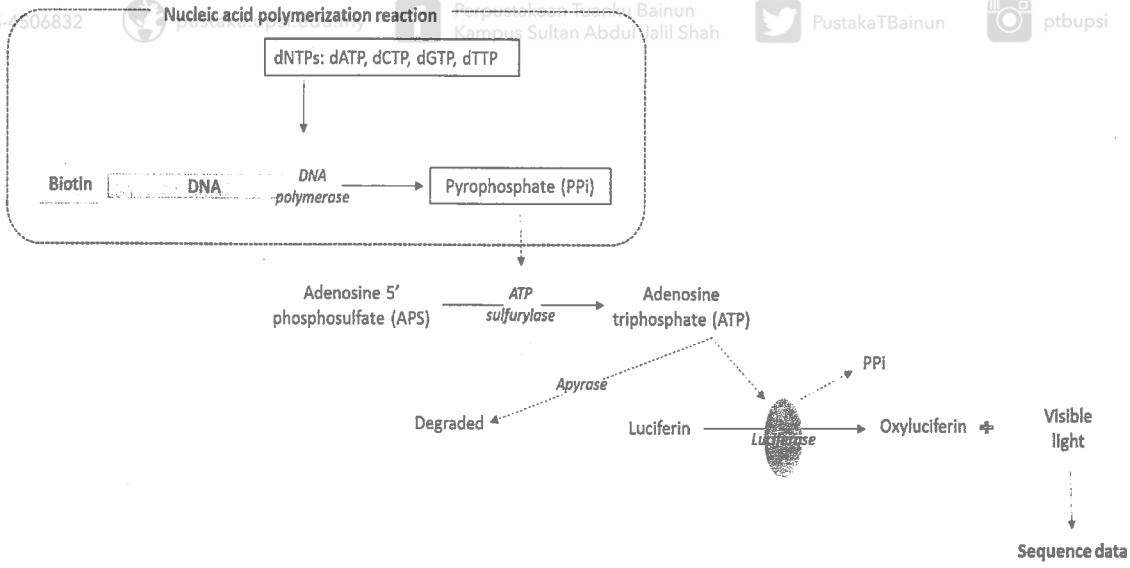

Figure 1.4 Schematic representation of pyrosequencing (Modified from Fakruddin et al., 2013).

1.6.2 The twin-track approach

Metagenomic analyses are being increasingly favoured over culture-dependent methods in surveys of microbial communities as they reveal the amazing levels of prokaryotic diversity, much of it novel, present in natural habitats. Indeed, culture-based methods are seen, with considerable justification, to be limited in scope as they are labour intensive, time-consuming, underestimate prokaryotic diversity and tend to overplay the significance of fast-growing and spore-forming taxa (Bull et al. 2000). However, it is also important to recognize that culture-independent data can be skewed by factors that include choice of PCR primers, differential cell lysis, rRNA copy numbers and the dominance of individual taxa at the time of sampling (O'Donnell et al. 1993; Bull et al. 2000; Engelbrektson et al. 2010) and by sequencing errors and data handling problems (Claesson et al. 2010; Kunin et al. 2010; Nakamura et al. 2011; Liu et al. 2012; Roh et al. 2010).

Given constraints such as those outlined above a case can be made for using both culture-dependent and culture-independent methods for determining the extent of prokaryotic diversity in natural bacterial communities. In such twin-track approaches the metagenomic element provides an invaluable perspective into the extent of prokaryotic diversity in environmental samples while corresponding culture-based analyses allow the selection and growth of dereplicated isolates for biotechnological and bioprospecting purposes. Within a broader context there is a need to devise innovative procedures to isolate culturable representatives of uncultured taxa for such purposes (Stewart 2012).

The twin-track approach though not widely applied has been used to analyse bacterial communities in anoxic rice paddy soil (Hengstmann et al. 1999), arid soils (Dunbar et al. 1999), the human gut (Wilson and Blitchington 1996; Suau et al. 1999), marine sediment (Chandler et al. 1997), seawater (Suzuki et al. 1997), in cold desert soils (Babalola et al. 2009), high temperature petroleum reservoirs (Orphan et al. 2000; Kaster et al. 2009) and more recently on glacier surfaces (Liu et al. 2015), in river and tapwater (Vaz-Moreira et al. 2011, 2013), and commercial salad leaf vegetables (Jackson et al. 2013). The outcome of such studies, while mixed, tends to show similar relationships between culture-dependent and culture-independent data though significant differences are also apparent. There are reports that culture-dependent and culture-independent methods target different bacteria (Vaz-Moreira et al. 2011) and that some taxa, including members of the rare biosphere, are detected only using culture-dependent procedures (Wilson and Blitchington 1996; Suau et al. 1999; Shade et al. 2012) .

1.6.3 Back to the Atacama Desert

Culture-independent methods have been used to establish the extent of prokaryotic diversity in Atacama Desert habitats and in some instances have been designed to gain an insight into the functional ecology of taxa (Azua-Bustos et al. 2012; Bull and Asenjo 2013). The outcome of such studies have been mixed as actinobacteria have been shown to be dominant in some habitats (Bull et al. 2016) but are absent or barely feature in others (Demergasso et al. 2010; Drees et al. 2004; de los Rios et al. 2010; DiRuggiero et al. 2013). The twin-track approach to establishing actinobacterial structure in Atacama Desert habitats has still to be realised.

Actinobacteria seem to predominate in soil samples taken from in and around the hyper-arid core region of the Atacama Desert, namely the Yungay region (Connon et al. 2007; Neilson et al. 2012; Crits-Christoph et al. 2013). Connon and her colleagues found that actinobacteria accounted for 94% of 16S rRNA genes cloned from Yungay soil samples, the majority of which were most closely, albeit very loosely related to the genera *Frankia* and *Rubrobacter* as well as to uncultured taxa. This dominance was underpinned in a 454-pyrosequencing study carried out by (Neilson et al. 2012) who found that members of the families *Conexibacteriaceae, Nutriliruptoraceae, Patulibacteriaceae* and *Solirubrobacteriaceae* and the subclass *Rubrobacteridae* were abundant in various Yungay samples, these taxa have rarely been isolated but are known to form deep-seated lineages in the actinobacterial 16S rRNA gene tree (Ludwig et al. 2012). Similarly, Crits-Christoph et al. (2013) analysed microbial communities at several geographical sites in

the Yungay region based on pyrosequencing of bar-coded 16S rRNA gene amplification products, generated using archeal and bacterial primer sets, and found communities dominated by actinobacteria belonging to the orders *Acidimicrobiales* and *Rubrobacteriales* and to representatives of the thermophilic families *Patulibacteriaceae* and *Solirubrobacteriaceae*. Sizeable actinobacterial populations have been found in hyper-arid, high altitude, non-fumerolic soil from Socompa Volcano on the Chilean-Argentine border (Costello et al. 2009) and in sediment samples from the Salar de Gorbea, south-east of Antofagasta (Bull and Asenjo 2013).

Functional ecological analyses of actinobacterial taxa found in Atacama Desert habitats are in their infancy. However, in one such study the metagenome generated from a high altitude debris field on Llullaillaco Volcano near the Chilean-Argentine border was almost solely composed of actinobacterial lineages, notably by the genome belonging to the genus *Pseudonocardia* (Lynch et al. 2012). Metabolic pathways encoded by this genome included an ability to use trace gases (CO, H_2) and C1 compounds, traits exhibited by several *Pseudonocardia* species (Huang and Goodfellow 2015)). It is also interesting that phylogenetic data from one of the sites studied by (Neilson et al. 2012) suggested the presence of chemoautotrophic taxa able to acquire energy through the oxidation of carbon monoxide, nitrate, iron and/or sulphur.