Molecular detection of type II polyketide synthase genes in Cuban soils.

By

Nathan Z. Morris, B.Sc. (Hons).

A thesis presented for the degree of
Doctor of Philosophy

Department of Biological Sciences
University of Warwick

September, 2000

Contents

Chapter 2 **Materials and Methods**

Summary

Molecular detection methods were developed to study the distribution of type II polyketide synthase (PKS) genes in Cuban soils. A PCR based detection method targeting the α and β ketosynthase genes was applied to a number of different total community DNA samples. These genes were detected in 43 % of samples tested from a number of different locations. A botanical garden site located in Havana, Cuba, was found to show the greatest distribution of type II PKS genes across the sites tested. It was not possible to amplify type II PKS genes from a pristine island site off the coast of Cuba. Further investigation revealed that actinomycetes containing type II PKS were present in the soil community at a level above the detection limit of the PCR protocol. Further total community DNA cleanup steps failed to allow the detection of type II PKS genes within the DNA samples suggesting PCR inhibition was responsible for negative results.

The molecular detection of type II PKS genes in total community DNA was compared to the detection of type II PKS genes in actinomycete isolates. A lack of correlation between these two approaches was observed with the molecular detection limit unable to amplify type II PKS genes in <50 % of crop soils tested. Actinomycetes containing type II PKS genes could be isolated from all crop soils tested. No difference was seen in the detection of type II PKS genes between rhizosphere and bulk soil samples. Actinomycetes were isolated using a selective isolation procedure at a level of approximately 10^7 cfu g^{-1} soil compared to 10^8 cfu g^{-1} for total bacterial counts.

Actinomycetes were isolated from Cuban crop soils and screened for the presence of type II PKS genes. Out of 100 isolates 26 were found to contain the genes of interest. Phylogenetic analysis of these isolates based on 16S rDNA and *recA* sequence data showed them to be closely grouped within the streptomycetes. Sequence data based on KSα genes from Cuban isolates showed them to be representative of both spore pigment and antibiotic polyketide genes.

A representative clone library was constructed containing type II PKS genes amplified from total community DNA. Rhizosphere and bulk soil samples were compared from the same site. Sequences obtained from rhizosphere total community DNA appeared to be widely distributed when compared to published sequences and included examples of both spore pigment and antibiotic polyketide genes.

A molecular method was developed to amplify near full length α and β KS genes from type II PKS gene clusters. Expression vectors were constructed to allow these genes to be expressed along with an ACP to give a functional minimal PKS for polyketide chain production. This method was used on total community DNA in an attempt to extract diverse genes from as yet uncultured organisms.

Abbreviations

aa	Amino acid
ATCC	American Type Culture Collection
ATP	Adenosine 5-triphosphate
bp	Base pairs
BSA	Bovine serum albumin
x g	Times gravitational units
cfu	Colony forming unit
°C	Degrees centigrade
d	Days
ddH_2O	Double distilled sterile water
DGGE	Denaturing gradient gel electrophoresis
DMSO	Di-methyl sulphoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DSM	Deutsche Sammlung von Mikroorganismen
e.g.	For example
EDTA	Ethylenediaminetetra-acetic acid
EtOH	Ethanol
g	Gram
G + C	Guanine and Cytosine
h	Hour
kb	Kilo-base pairs
ks	ketosynthase
LB	Luria broth
µ	Micro
µg	Micrograms
µg ml^{-1}	Micrograms per millilitre
mg	Milligrams
mg ml^{-1}	Milligrams per millilitre
min	Minute

ml	Millilitre
mM	Millimolar
M	Mole
NA	Nutrient agar
ng	Nanograms
no.	Number
nt.	Nucleotide
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
%	Percent
PKS	Polyketide synthase
PSI	Pounds per square inch
PVPP	Polyvinylpolypyrolidine
RASS	Reduced arginine salts solution
RNA	Ribonucleic acid
Rnase	Ribonuclease
rpm	Revolutions per minute.
rRNA	Ribosomal ribonucleic acid
s	Second
SDS	Sodium dodecyl sulphate
SDW	Sterile distilled water
spp.	Species
TAE	Tris-acetate EDTA
TBE	Tris borate EDTA
TE	Tris-EDTA
TRIS	Tris (hydroxymethyl) aminomethane
UV	Ultraviolet
V	Volts
v / v	volume / volume
w / v	weight / volume

List of Tables

List of Figures

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results were obtained by myself and all sources of information have been specifically acknowledged by source of reference.

Nathan Z. Morris

Acknowledgements

I would like to thank my supervisor Professor Liz Wellington, for all her help over the duration of this project. I would especially like to thank her for the encouragement and enthusiasm she has shown towards the project and myself.

I gratefully acknowledge Novo Nordisk and the BBSRC for their financial support of this project and for the provision of grants awarded for overseas travel.

I would like to thank Professor C.R. Hutchinson and members of his laboratory at the University of Wisconsin, Madison, for enabling me to spend some time working in their lab (July – September, 1999).

My thanks also go to everyone who has helped me during my time at Warwick especially Jane for her assistance in the lab.

Publications and Presentations

Krsek, M., **Morris, N.Z.,** Egan, S. and Wellington, E.M.H. (1999). Actinomycetes. In *Encyclopedia of Microbiology*. 2nd edn. Academic Press.

Morris, N.Z., Standage, S. and Wellington, E.M.H. (1998). DNA extraction from spores in soil. In *Molecular Ecology Manual*. 4th edn, 1.3.7. Edited by A.D.L. Akkermans, J.D. van Elsas and F.J. de Bruijn. Kluwer Academic Publishers.

Morris, N.Z., Gürtler, H. and Wellington, E.M.H. (1999). Molecular detection of type II polyketide synthase genes in Cuban soils. *11th International Symposium on the Biology of Actinomyctes (ISBA)*. Crete, Greece, October 1999 (Poster).

Morris, N.Z., Gürtler, H. and Wellington, E.M.H. (1999). Molecular detection of type II polyketide synthase genes in Cuban soils. *SGM*. Edinburough, UK, April 1999 (Poster).

Morris, N.Z., Gürtler, H. and Wellington, E.M.H. (1999). Molecular detection of type II polyketide synthase genes in Cuban soils. *Molecular Microbial Ecology 5*, Aberdeen, UK, April 1999 (Oral).

Morris, N.Z., Revill, P., Gürtler, H. and Wellington, E.M.H. Distribution and diversity of type II polyketide synthase genes in soil populations of actinomycetes colonising Cuban soils. *Portrait of an Organism: The Genetic Analysis of Streptomyces coelicolor A3 (2) Biology*. 141st ordinary meeting of the SGM, Norwich, UK, September 1998 (Poster).

Morris, N.Z., Revill, P., Gürtler, H. and Wellington, E.M.H. Distribution and diversity of type II polyketide synthase genes in soil populations of actinomycetes colonising Cuban soils. *Polyketides II, Chemistry, Biochemistry and Molecular Genetics*, Bristol, UK, July 1998 (Poster).

Morris, N.Z., Revill, P., Gürtler, H. and Wellington, E.M.H. Distribution and diversity of type II polyketide synthase genes in soil populations of actinomycetes colonising Cuban soils. *8th International Symposium on the Genetics of Industrial Microorganisms (GIM98)*, Jerusalem, Israel, June 1998 (Poster).

Wellington, E.M.H., Huddleston, A.S., Krsek, M., Egan, S., Standage, S. and **Morris N.Z.** DNA extraction and detection of antibiotic biosynthesis genes in nature. *Rhone-Poulenc - How best to generate leads from natural substances*, Paris, France, October 1997.

Morris, N.Z., Revill, P., Gürtler, H. and Wellington, E.M.H. Distribution and diversity of type II polyketide synthase genes in soil populations of actinomycetes colonising Cuban soils. *10th International Symposium on the Biology of Actinomycetes (ISBA)*, Beijing, China, May 1997 (Poster).

Chapter 1

General Introduction

1.1. The Actinomycetes

The actinomycetes are Gram-positive bacteria which have a characteristically high G + C content in their DNA (>55 %). These bacteria are significant because they produce a large number of biologically active compounds, include major pathogens and play important roles in the environment (Williams, 1982). Many species are of commercial importance due to the bioactive secondary metabolites that they produce (Strohl, 1997). The actinomycetes were originally considered to be an intermediate group between bacteria and fungi but are now recognised as prokaryotic. They are also among the most morphologically complex of bacteria. Depending on the taxon, they may produce branched rods, complicated mycelial structures, spore bodies, and motile or nonmotile spores (Embley and Stackebrandt, 1994). Over 80 genera have now been described and the most widely distributed are members of the genera *Micromonospora*, *Nocardia*, *Rhodococcus* and *Streptomyces* (Wellington and Toth, 1994). Traditional methods based upon morphological, phenotypic and chemotaxonomic characteristics have been used to classify actinomycetes. These characteristics have shown varying degrees of coherency with modern phylogenetic classification methods (Stackebrandt *et al.*, 1983). Each of these approaches though has provided important information on taxonomic structure, and represents part of a polyphasic approach for the identification of actinomycetes at the genus and species levels (Stackebrandt and Liesack, 1993).

1.2. Classification of actinomycetes

Actinomycete classification has relied heavily on morphological characterisation. This was summarised by Ensign (1992) who used morphological diversity as a method to define phylogenetic groups. They included genera forming cocci (*Kineococcus* and *Micrococcus*), pleomorphic rods (*Arthrobacter* and *Corynebacterium*) and those with branching, filamentous mycelium, such as most of the sporoactinomycetes. The later group form spores on the aerial and / or substrate mycelium either singly (*Micromonospora* and *Thermomonospora*) or in chains (*Actinomadura*, *Pseudonocardia* and *Streptomyces*). Some genera also produce sporangia (*Actinoplanes*, *Frankia* and *Streptosporangium*), which may contain motile spores. It has been shown at the genus level that morphological based classification is inadequate when compared to numerical taxonomy (Williams *et al.*, 1983) and phylogenetic analyses of streptomycetes (Witt and Stackebrandt, 1990; Wellington *et al.*, 1992). Phenotypic data has often been used to identify microorganisms but can be treated more objectively using numerical taxonomic procedures (Sneath, 1989). The use of numerical taxonomy and its application in ecological studies has been reviewed by Goodfellow and Dickinson (1985). Molecular tools have been used for classifying actinomycetes and this can result in actinomycetes being clustered together even when they have different biochemical traits. Consequently identifications made to the genus level based on biochemical characteristics may differ from the identification predicted by molecular information from sequencing. Sequence information from the 16S rDNA genes can be used to develop specific hybridisation probes for identifying microorganisms. PCR primers based on 16S rDNA genes coupled with sequencing can be used to

classify actinomycetes at the genus and species level (Stackebrandt *et al.*, 1991). The actinomycetes as defined by 16S rDNA sequence analysis form a tight clade in the prokaryotic phylogenetic tree (Figure 1.1). Partial sequencing of variable regions of 16S ribosomal RNA has allowed intergeneric classification of actinomycetes (Figure 1.2).

1.3. Ecology of actinomycetes

Actinomycetes are ubiquitous in soil and are also found in plant litter, compost, freshwater, salt water, and can colonise plants and animals. They are able to utilise a wide range of substrates and can undergo parasitic or mutualistic relationships with plants and animals. Soil and aquatic environments are reservoirs for high diversity of prokaryotes and over 20 actinomycete genera have been isolated from soil (Williams and Wellington, 1982). Viable counts have been recorded in the region of 10^6 cfu g^{-1} in fertile soil (Goodfellow and Williams, 1982). In aquatic systems the highest actinomycete diversity is found in shallow waters and sediment. The predominant genera cultured in these environments include *Streptomyces*, *Micromonospora*, *Microbispora*, *Nocardia* and *Thermoactinomyces* (Cross and Johnston, 1972; Weyland, 1981). Molecular techniques can be used to reveal details of uncultured actinomycetes in the natural environment. Felske *et al.* (1997) used ribosomal analysis to study the activity of uncultured actinomycetes in grassland soils.

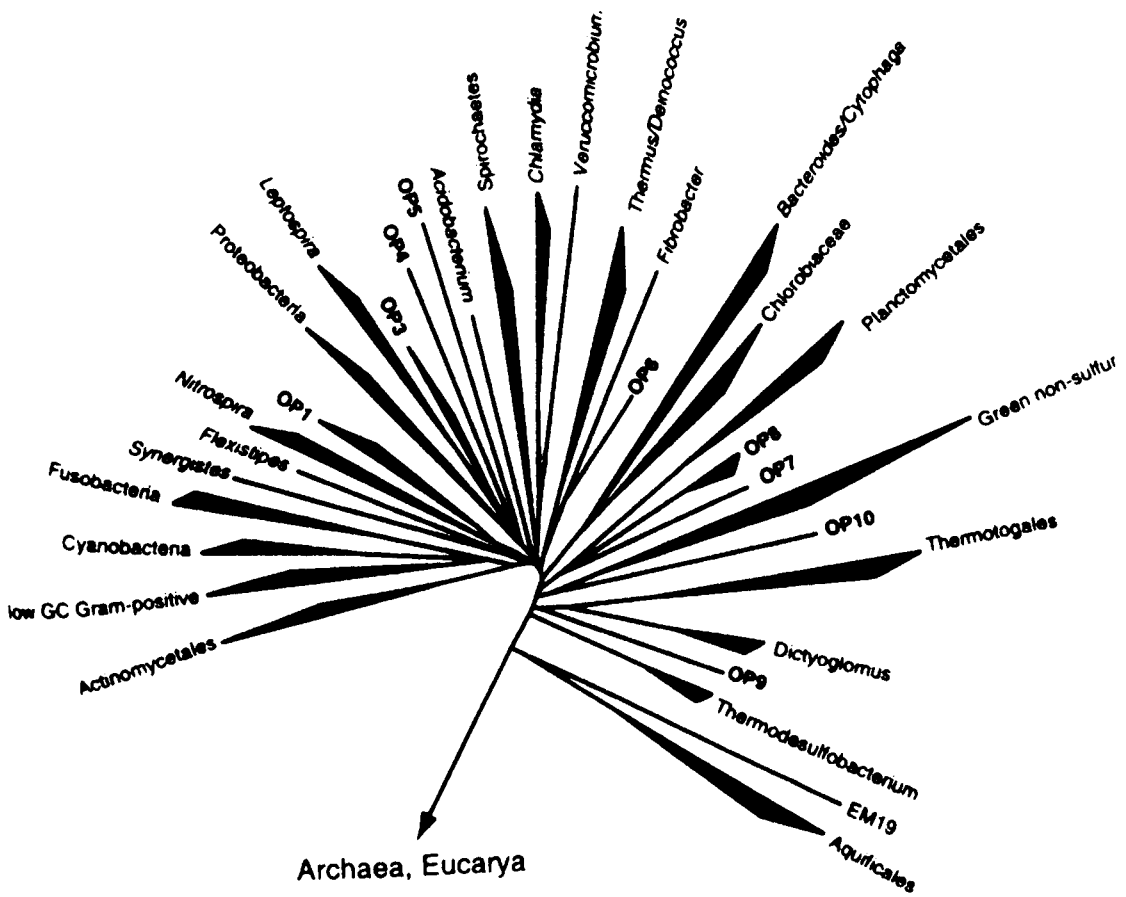


Figure 1.1. Diagrammatic representation of the known phylogenetic span of bacteria based on 16S DNA sequence analysis (reproduced from Pace, 1997).

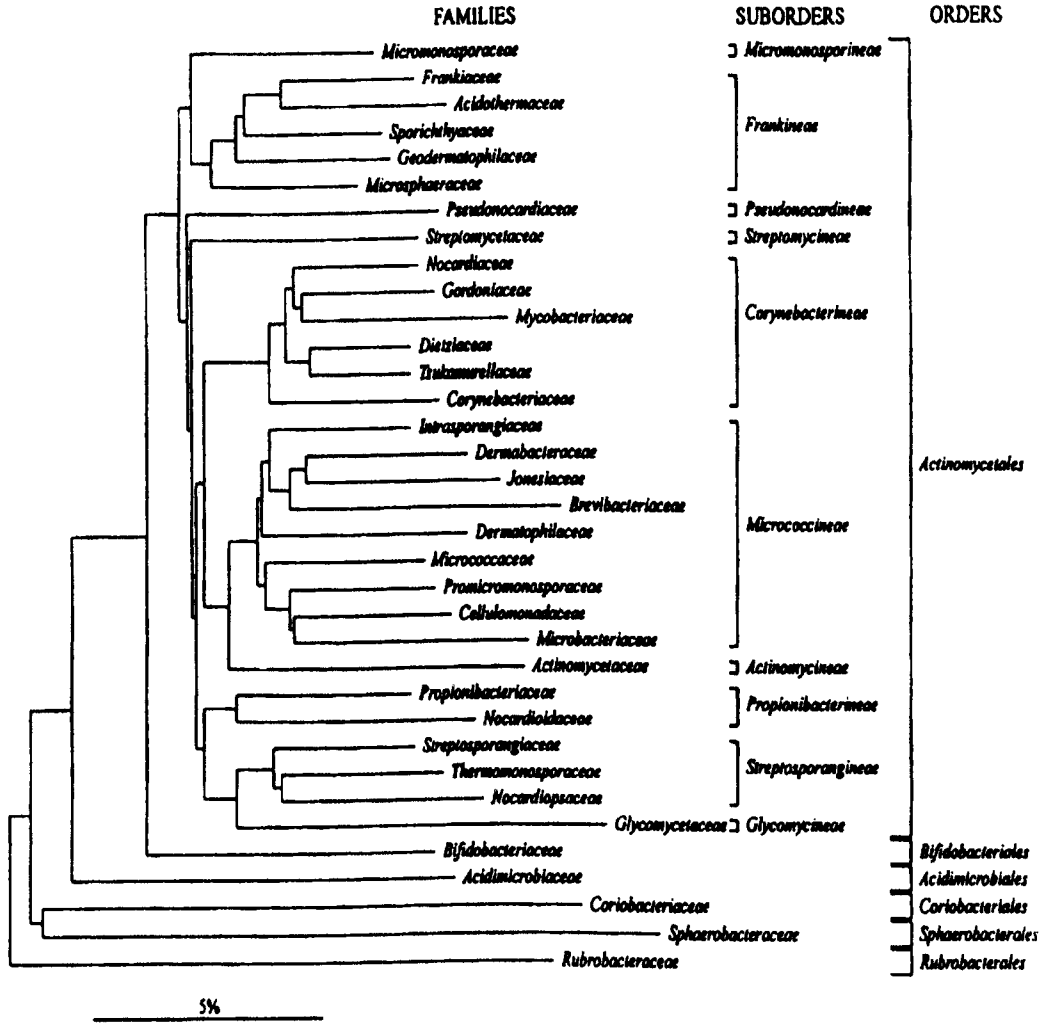


Figure 1.2 Interclass relatedness of actinobacteria based upon 16S rDNA/rRNA sequence comparison. The scale bar represents 5 nucleotide substitutions per 100 nucleotides (Reproduced from Stackebrandt *et al.*, 1997).

1.4. The terrestrial environment

Soil is a complex, constantly changing environment which, due to its high solid / liquid ratio contains many diverse microhabitats for microorganisms. Soil formation is the result of a number of interactive forces including microbial activity, climate, topography and parent material (Paul and Clark, 1989). Organic matter comprises 3 to 6 % of the total volume of soil, consisting of decomposing plants, animals, microorganisms and humus (humic acid, fulvic acid and humin) which represents incomplete decomposed organic material. Humus is formed from compounds such a cellulose and chitin which are resistant to decomposition (Gray and Williams, 1971).

Carbon and energy sources
Mineral nutrients
Growth factors
Ionic composition
Water availability
Temperature
Atmospheric composition
pH
Oxidation-reduction potential
Genetics of the microorganisms
Interactions between microorganisms

Table 1.1. Factors affecting the activity, ecology, and population dynamics of microorganisms in natural habitats (Stotzky, 1997).

Water found within the soil environment forms a liquid medium for microorganisms and availability depends on a number of factors. These include water volume, soil porosity and the degree of water adsorption by clay particles. In dry soils pore spaces are filled with air with carbon dioxide concentrations generally increasing with depth. This may not apply though to specific niches due to the complex heterogeneity of soil structure (Paul and Clark, 1989). Some of the

physiochemical characteristics of soil that have been demonstrated to affect the activity, ecology and population dynamics of microorganisms are listed in Table 1.1.

Clay minerals present in soil have been shown to have a profound influence on microbial activities in soil. These include growth, spore germination, pathogenesis and transfer of genetic information as well as providing protection against toxicity, elevated temperatures, desiccation and ultraviolet light. How clay minerals affect these activities is not always clear. In some cases the affects of clays appears to be indirect by modifying the physiochemical characteristics of the microhabitats, thereby either enhancing or attenuating the growth or metabolic activity of the microbial population (Stotzky, 1992, Figure 1.3).

1.5. The Rhizosphere

Hiltner (1904; quoted in Stotzky, 1997) first introduced the term "rhizosphere", defining it as "that volume of soil surrounding roots in which bacterial growth is stimulated". The term "endorhizosphere" may now be used to describe the microenvironment surrounding the plant root (Belandreau and Knowles, 1978). This can include microbe-derived polysaccharide and the epidermal layers of plants. The rhizoplane can be defined as the epidermal layer, including its associated polysaccharide matrix. The ectorhizosphere comprises of soil extending a few millimetres from the root system. Bulk soil can be separated from rhizosphere soil by shaking or washing the plant roots. Soil adhering to the root can be defined as the rhizosphere. The rhizosphere is a dynamic environment and production and

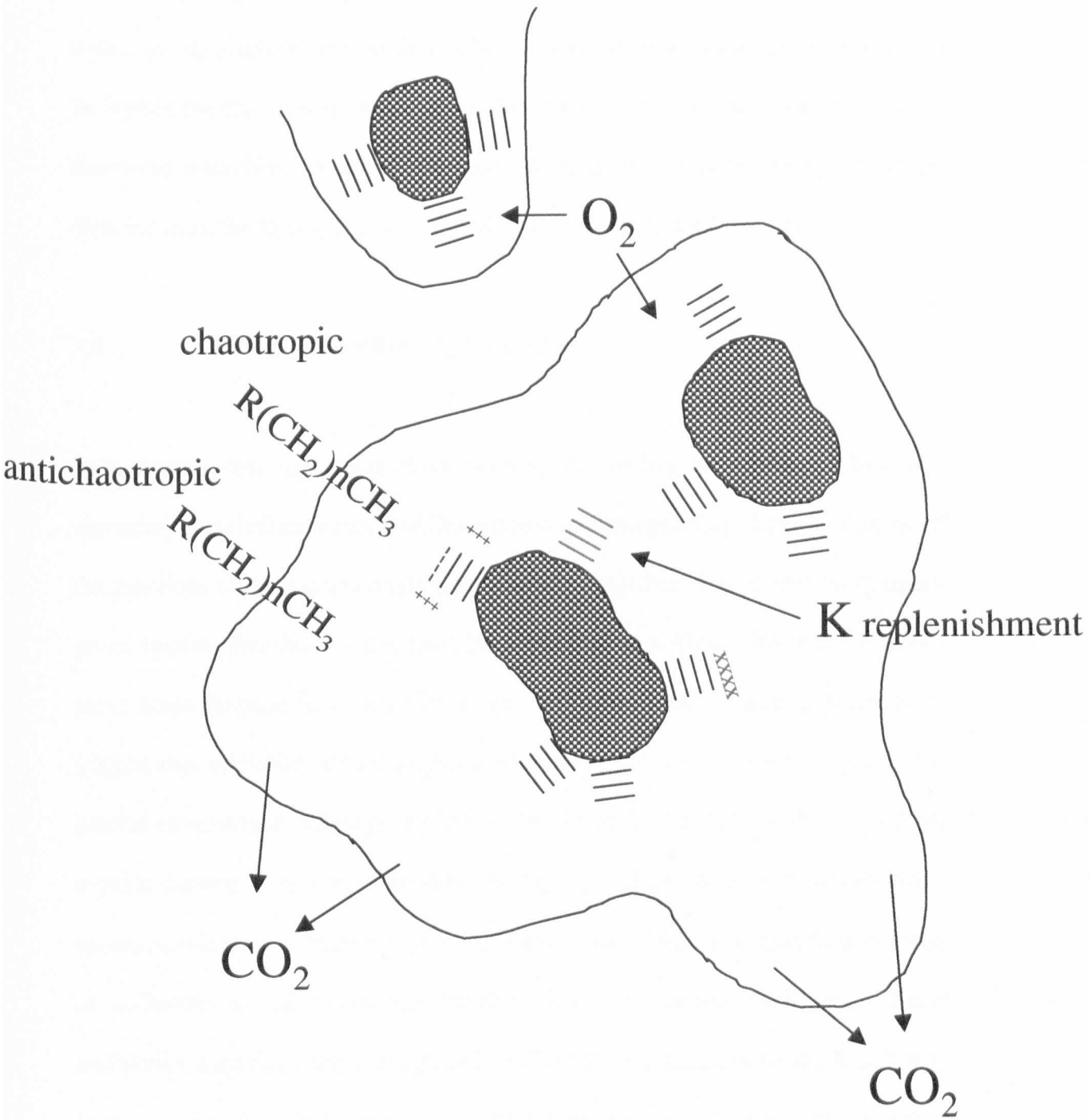


Figure 1.3. Schematic representation within a microhabitat in soil. The patterned areas represent sand or silt particles, and the families of adjacent short lines represent packets of clay minerals, with the location of positive and negative charges indicated. The arrows represent movement of O_2, CO_2 and nutrients and XXXX indicated bound toxic substances. Adapted from Stotzky, 1997.

exudation of organic compounds from plant cells affects the microbial community within the rhizosphere. Due to these plant exudates microbial activity is expected to be high in the rhizosphere. Højberg and Sørenson (1993) used O_2 concentrations to determine microbial activity in the rhizosphere of gel-stabilised barley. Increased distance from the rhizosphere was found to decrease microbial activity.

1.6. Ecology of antibiotic production

It is argued that organisms have evolved the ability to synthesise bioactive secondary metabolites because of the selective advantages they obtain as a result of the functions of these compounds. The clustering together of antibiotic biosynthetic genes implies that these genes have been selected as a group and that antibiotics serve some purpose in nature (Stone and Williams, 1992). There is evidence to suggest that microbial natural products with biological activity are produced in the natural environment although antibiotics are difficult to detect *in situ*. In soil the organic content is very low resulting in only a small percentage of streptomycete spores germinating to form mycelia (Cresswell *et al.*, 1992). This may lead to yields of antibiotics in soil below the detection limits for antibiotic extraction. Some antibiotics, especially basic compounds, adhere to clay particles in soil which may lead to extraction difficulties. Antibiotics may also be degraded by microbial activity in soil (Williams, 1982). Several studies have shown production of metabolites in soil. *S. venezuelae* was found to produce chloramphenicol in sterile soil microcosms (Gottlieb, 1976). A direct correlation has also been observed between the level of antibiotic production by a bacterial inoculant and the nutrient levels present in sterile soil microcosms (Soulides, 1965). Antibiosis has also been

observed in the rhizosphere including geldanomycin producing *S. hygroscopicus* var. *geldanus* preventing root rot by *Rhizoctania solani* in the rhizosphere of pea crops (Rothrock *et al.*, 1984).

1.7. Antibiotics

Antibiotic production by bacteria is one of their most interesting characteristics, particularly from the viewpoint of the medical and commercial value of these compounds (Hutchinson *et al.*, 1993). Waksman coined the term "antibiotic" and described an antibiotic as "a chemical substance derived from microorganisms which has the capacity of inhibiting growth, and even destroying, other microorganisms in dilute solutions" (Baltz, 1986).

Antibiological activity	**Example of natural products**
Antibacterial	Streptomycin, tetracycline
Antifungal	Amphotericin B, nystatin
Antiviral	Lamivudine (3TC)
Antihelminthic	Ivermectin
Bacteriostatic cationic peptides	Magainins, defensins, sapecins
Insecticidal	*Bacillus thuringiensis* σ-toxin, Spinosad
Anticoccidial	Monensin, nosiheptide
Antitumor	Doxorubicin, mitomycin C
Immunosuppressive	Cyclosporin, rapamycin
Growth promotors	Monensin
Anticholesterolemic	Lovastatin, pravistatin

Table 1.2. Antibiological activities of natural products in use, in development, or of potential future use (Reproduced from Strohl, 1997).

Natural product antibiotics belong to a group of compounds called secondary metabolites. These are characterised by having structures that are unusual compared with those of intermediary metabolites, by being produced at low specific growth

rates, and by the fact that they are not essential for growth of the producing organisms in pure culture (Queener and Lively, 1986).

Antibiotics though, are believed to be crucial to the producing organisms in the natural environment. Reasons for this include survival and competitive advantage (Hopwood *et al.*, 1985). Although Waksman's definition of antibiotics is still true there are now a vast assortment of antibacterial, antifungal, antiviral, antitumor and antimetabolic substances known to be produced by both micro- and macroorganisms today. The world market for these anti-infectives has now surged from $18 billion in 1994 to $23 billion in 1995 (Strohl, 1997).

1.8. Antibiotics produced by Actinomycetes

Actinomycetes in particular members of the *Streptomyces* genus are of industrial importance due to the wide variety of biologically active secondary metabolites that they produce including antibiotics (Wellington and Cross, 1983). The global market for microbial derived pharmaceuticals was worth $28 billion in 1996, luring the pharmaceutical industry into funding research in the hope of commercial profit (Strohl, 1997). These include compounds which many different activities. Some strains even produce more than one kind of compound not even chemically related. These compounds are classified into chemical families including aminoglycosides, anthracyclines, β-lactams, macrolides, peptides and tetracyclines (Okami and Hotta, 1998). These compounds vary in structure from simple sugars (e.g. aminoglycosides) to complex macromolecules (e.g. tetracyclines). Also different strains can produce the same antibiotic. For example erythromycin B is produced by

Micromonospora sp. 1225 and was the first example of an erythromycin to be produced outside *Saccharopolyspora* spp. (Marquez *et al.*, 1976).

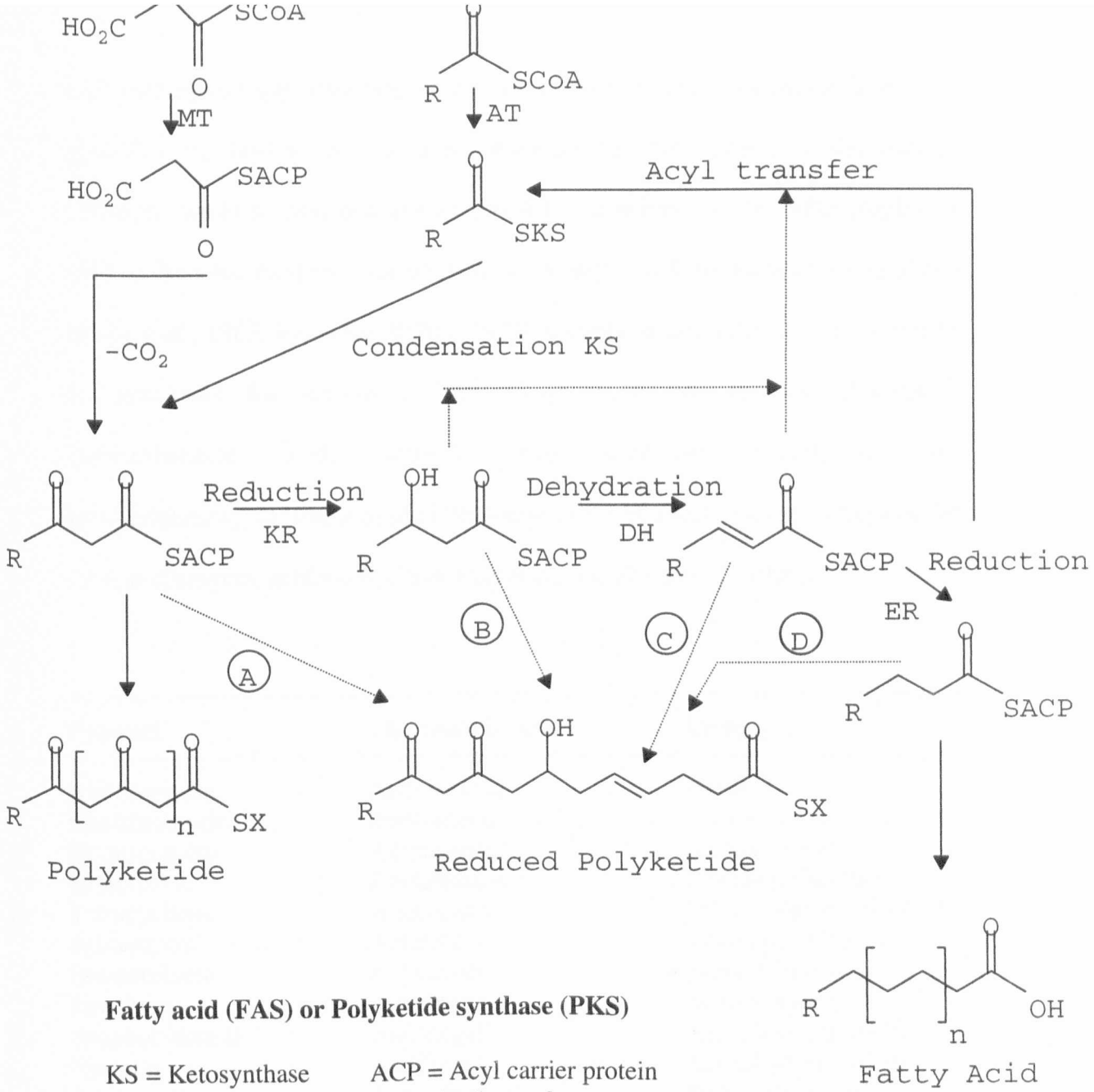
1.9. Polyketides

The polyketide metabolites are a large group of structurally complex and diverse natural products, many of which are clinically valuable antibiotics, chemotherapeutic agents or have other useful pharmacological activities (Monaghan and Tkacz, 1990). One of the smallest polyketides, 6-methylsalicyclic acid, is a single aromatic ring (Beck *et al.*, 1990), whereas one of the largest polyketides, Brevitoxin B, contains 50 carbon atoms in its chain (Katz and Donadio, 1993). The biological activities of polyketides have been known indirectly since ancient times. Purgative materials in cascara, rhubarb and senna are usually polyketide-derived anthracenes. Traditionally antispasmodics, used in the Middle East to treat angina, contain significant amounts of the polyketide khellin (5,8-dimethoxy-2-methyl-6,7-furanochromone). One of the most famous uses of a polyketide in history was the use of coniine-containing hemlock to execute Socrates (399 BC) by lethal ingestion (Bentley and Bennett, 1999). Polyketides can also be multifunctional. A well characterised example is the antibiotic erythromycin, produced by *Saccharopolyspora erythraea*. Its major clinical use is as a protein synthesis inhibitor when used to treat Gram positive bacterial infections (Donadio *et al.*, 1992). It has been used for this clinically for 50 years but the compound also induces gastric contractions (Itoh *et al.*, 1984) and can act as a motilin agonist (Omura, 1992).

Despite their apparent structural diversity, polyketides share a common mechanism of biosynthesis. As first suggested by Colie in 1907, the chemical synthesis of polyketides is centred around the reactive groups of the ketone and its adjacent α-methylene carbon (Colie, 1907; quoted in Kramer and Khosla, 1996). Therefore 3 opportunities exist for structural diversity, the nature of the starter unit, the extender unit and the stereochemistry resulting from their condensation. The carbon backbone of a polyketide results from sequential condensation of short chain, aliphatic acids, such as acetate, propionate, or butyrate, in a manner resembling fatty acid biosynthesis but catalysed by polyketide synthases (PKSs, Figure 1.4; Hopwood, 1997).

Unlike fatty acid biosynthesis, the carbonyl groups of the growing polyketide chain are not always fully reduced during biosynthesis, leaving a reactive 3-oxoacyl thioester intermediate, presumably still attached to the enzyme, that can be elaborated in a variety of ways (Summers *et al.*, 1995). A strong sequence and mechanistic similarity among many of the PKSs has led to two theories for explaining polyketide biochemistry: type I PKSs, multifunctional proteins that harbour a distinct active site for every enzyme-catalysed step, like macrolide antibiotics, and type II PKSs, multienzyme complexes that carry a single set of iteratively used activities and consist of several largely monofunctional proteins for the synthesis of complex largely reduced polyketides like aromatic polyketides, for example tetracyclines (Katz and Donadio, 1993; Hutchinson and Fujii, 1995).

The shikimate pathway and variations of this pathway play a crucial role in providing aromatic precursors for the biosynthesis of natural products particularly



Fatty acid (FAS) or Polyketide synthase (PKS)

KS = Ketosynthase
KR = Ketoreductase
DH = Dehydratase
ER = Enoyl reductase

ACP = Acyl carrier protein
AT = Acyl transferase
MT = Malonyl transferase

Figure 1.4. The basic pathway of fatty acid or polyketide biosynthesis, showing the roles of the various activities carried out by the subunits or domains of the FAS or PKS.
Alternate versions of the reductive cycle that lead to the keto (A), hydroxy (B), enoyl (C), or methylene (D) functionality during assembly of reduced polyketides (Hopwood, 1997)

in *Streptomyces* spp. (Wallace *et al.*, 1994). For instance, shikimate-derived *p*-aminobenzoic acid is used as a precursor for the biosynthesis of the polyene antibiotic candicin (Martin and Liras, 1976), and shikimate derived tryptophan is used as the aromatic precursor for both actinomycin and streptonigrin biosynthesis (Herlt *et al.*, 1985; Jones and Keller, 1997). Closely related pathways are also used as precursors for natural product biosynthesis. For example 3-amino-5-hydroxybenzoic acid, derived from erythrose 4-phosphate and phosphophenolpyruvate, is used in the biosynthesis of a wide variety of polyketides such as rifamycin, geldanomycin and mitomycin C (Kim *et al.*, 1996).

Product	**Therapeutic Area**	**Company**
Azithromycin	Antibacterial	Pfizer
Clarithromycin	Antibacterial	Abbott
Erythromycin	Antibacterial	Abbott, others
Rifamycins	Antibacterial	Novartis, Lepetit
Tetracyclines	Antibacterial	Pfizer, Wyeth-Ayerst
Adriamycin	Anticancer	Pharmacia-Upjohn
Daunorubicin	Anticancer	Astra, Chiron
Enediynes	Anticancer	Wyeth-Ayerst
Amphotericin B	Antifungal	Bristol-Myers Squib
Nystatin	Antifungal	Bristol-Myers Squib
Spiramycin	Antirickettsial	Rhône-Poulenc
Mevacor (Lovastatin)	Cholesterol-lowering	Merck
Pravastatin	Cholesterol-lowering	Bristol-Myers Squib
Zocor	Cholesterol-lowering	Merck
FK506	Immunosuppressant	Fujisawa
Avermectin	Veterinary product	Merck
Monensin	Veterinary product	Lilly
Tylosin	Veterinary product	Lilly

Table 1.3. Major pharmaceutical products derived from polyketides. Market data available ($, year) for some of these products are as follows: Azithromycin, Clarithromycin and Erythromycin ($3500 M, 1996), Amphotericin B and Nystatin ($300 M, 1991), Mevacor, Pravastatin and Zocor ($5200 M, 1996) and Tetracyclines ($664 M, 1996). Adapted from the Kosan Biosciences, Inc. Scientific Prospectus (http://www.kosan.com/K3prospectus.html).

1.10. The genetics of polyketide synthases (PKSs)

The analogy between polyketide and long-chain fatty acid biosynthesis has been extended to the genetic level by studies that have demonstrated similarity between the products of several PKS genes (Hopwood and Sherman, 1990; Katz and Donadio, 1993) and their fatty acid synthase congeners in *Escherichia coli* (Vanden Boom and Cronan, 1989; Magnuson *et al.*, 1993), yeast (Schweizer *et al.*, 1987) and mammals (Witkowski *et al.*, 1991). Examples have been found of PKSs resembling each of the classical classes of fatty acid synthases (FASs). The PKSs are classified in the same way as FASs of which the type I and II are found mainly in bacteria and fungi (Hutchinson *et al.*, 1993). It is believed that the type I and II PKSs share the same evolutionary origin but their sequences are too far diverged for significant DNA hybridisation (McCarthy and Hardie, 1984). The first example of a fungal PKS had a type I organisation while the first bacterial PKS to be studied, for members of the aromatic family of polyketides from actinomycetes, turned out to have a type II structure. Unexpectedly the gene sequences for PKSs for the macrolide polyketides of actinomycetes turned out to have not only a type I organisation previously only seen in eukaryotes, but the presence of multiple sets, or modules of active sites (Hopwood, 1997).

1.10.1. Type I PKSs

In 1982 a DNA fragment was isolated that conferred resistance to the macrolide antibiotic erythromycin on *S. lividans* (Thompson *et al*, 1982). This resistance determinant, *ermE,* was sequenced and used as a probe for the erythromycin

biosynthetic genes in *Saccharopolyspora erythraea* in the laboratories of R.H. Baltz, L. Katz, C.R. Hutchinson and P.F. Leadley. One *eryA* mutant, which was blocked in the formation of the aglycon moiety of erythromycin (6-deoxyerythronolide B; Hallam *et al.*, 1988) was found to be complemented by a piece of DNA located approximately 12 kb downstream of *eryE* (Haydock *et al.*, 1991). More significantly another segment of DNA that hybridized and complemented an *eryA* mutant was found 35 kb downstream of *ermE* suggesting that the *eryA* locus was very large (Tuan e*t al.*, 1990). The sequencing of this *eryA* region in the laboratories of L. Katz and P.F. Leadley revealed 3 large genes each encoding a protein carrying 2 modules of PKS active sites. Each module resembled in both sequence and organisation a vertebrate fatty acid synthase. These proteins known as 6-deoxyerythronolide B synthase (DEBS) 1, 2, and 3 (Figure 1.5) are arranged in the same order as their sequence of action (Cortes *et al.*, 1990; Donadio *et al*; 1991; Bevitt *et al*; 1992).

Other type I PKS systems have now been studied such as those responsible for the biosynthesis of the polyketide precursors of rapamycin (Schwecke *et al.*, 1995), FK506 (Motamedi e*t al.*, in press) and rifamycin (Schupp *et al.*, 1998) leading to certain common observations. Modular PKSs are organised into groups of active sites known as modules with each module being responsible for one cycle of polyketide chain extension. Within each module are catalytic domains of 100-400 amino acids each that are analogous in both function and sequences to the individual enzymes of fatty acid biosynthesis. All modules possess at a minimum β-keto acylthioester synthase [ketosynthase (KS)], acyl transferase (AT) and acyl carrier protein (ACP) domains. In addition, specific combinations of keto reductase

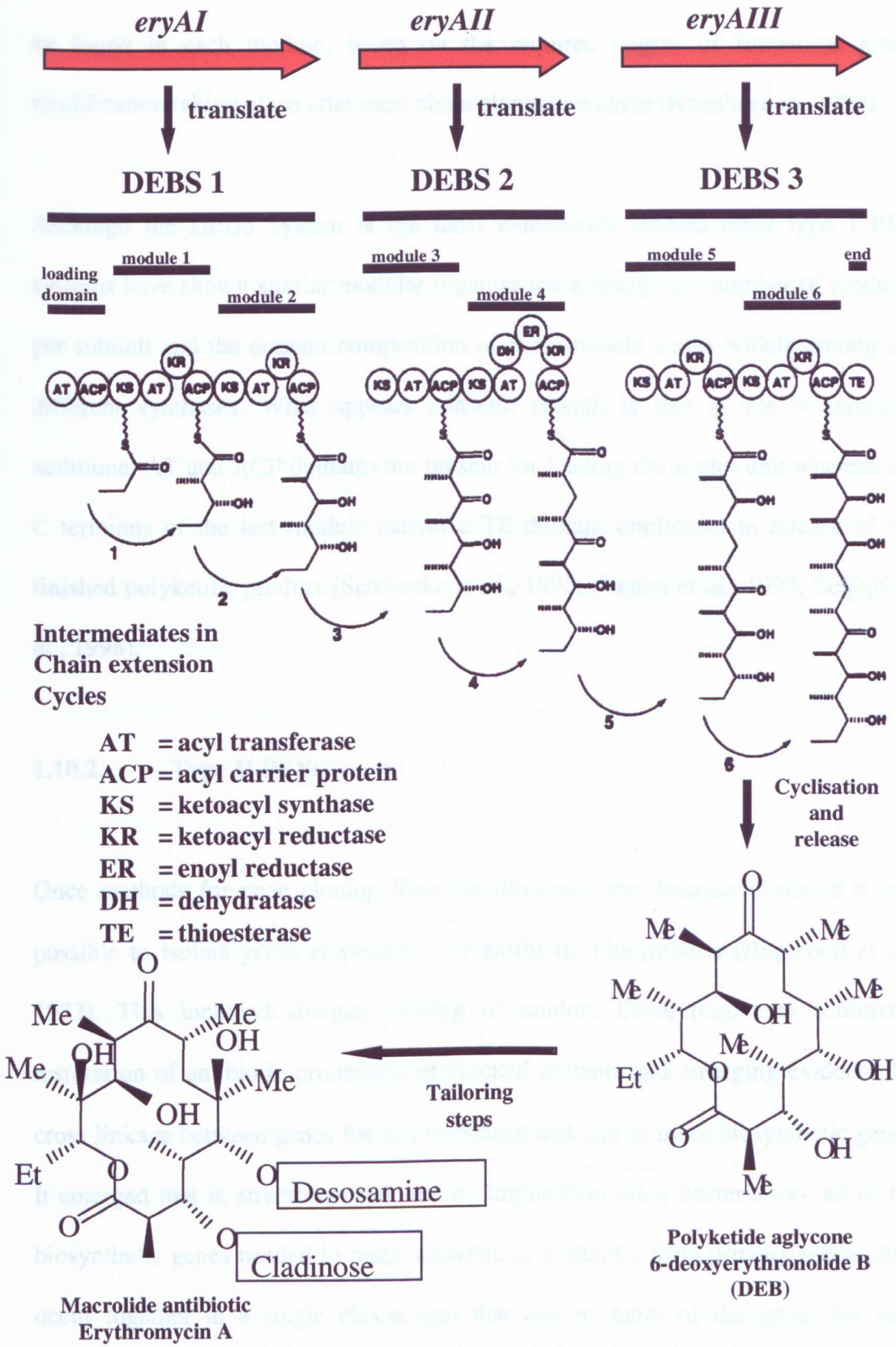


Figure 1.5. The erythromycin PKS (DEBS) consists of three giant multifunctional proteins forming the polyketide aglycone, 6-deoxyerythronolide B and on to erythromycin A (reproduced from Staunton and Wilkinson, 1997; Hopwood 1997).

(KR), dehydratase (DH), enoyl reductase (ER) and thioesterase (TE) domains may be found in each module, based on the required degree of functional group modification taking place after each chain elongation cycle (Khosla *et al.*, 1999).

Although the DEBS system is the most extensively studied other type I PKS systems have shown similar modular organisation although the number of modules per subunit and the domain composition of each module varies widely among the different synthases. What appears constant though is that at the N terminus additional AT and ACP domains are present for loading the starter unit whereas the C terminus of the last module carries a TE domain implicated in release of the finished polyketide product (Schweeke *et al.*, 1995; August *et al.*, 1998; Schupp *et al.*, 1998).

1.10.2. Type II PKSs

Once methods for gene cloning from *Streptomyces* spp. became available it was possible to isolate genes responsible for antibiotic biosynthesis (Hopwood *et al.*, 1983). This included shotgun cloning of random DNA fragments conferring restoration of antibiotic production in blocked mutants and emerging evidence for cross linkage between genes for self resistance and one or more biosynthetic genes. It emerged that in streptomycetes and by implication other bacteria too, all of the biosynthetic genes needed to make a particular antibiotic from primary metabolites occur together in a single cluster and that one or more of the genes for self-resistance are also present (Chater and Bruton, 1985). The first example of this in a type II PKS was the cloning of the entire cluster of genes (the *act* cluster) for

biosynthesis of the pigmented benzoisochromanequinone polyketide actinorhodin on a 35 kb fragment of chromosomal DNA from the producer *S. coelicolor* A3(2). Expression of this fragment in the actinorhodin-sensitive and nonproducing strain, *S. parvulus*, gave rise to actinorhodin without killing the host (Malpartida and Hopwood, 1984).

This led to a general approach for isolating antibiotic biosynthetic genes responsible for the biosynthesis of polyketides by actinomycetes:

1. Cloning of libraries of large fragments of wild-type DNA into available mutants blocked in steps required for antibiotic biosynthesis, looking for complementation, and then finding genes for other steps on the pathway on the complementing fragments.
2. Cloning a library of DNA fragments from an antibiotic producer into a sensitive surrogate host, selecting resistant clones and looking for biosynthetic genes linked to the resistance gene on the cloned DNA (Hopwood, 1997).

These methods allowed the biosynthetic genes for two more type II polyketides to be isolated including the anthracycline antibiotic, tetracenomycin (*tcm*), from *S. glaucescens* (Motamedi and Hutchinson, 1987), and oxytetracycline (*otc*) from *S. rimosus* (Butler *et al.*, 1989).

Areas that were believed to encode the KS and KR functions within the actinorhodin cluster (*actI* and *actIII*; Malpartida and Hopwood, 1986) had been deduced and these could be used as hybridisation probes against restriction digests

of *tcm* and *otc* cloned DNA (Malpartida *et al.*, 1987). In the same study these *actI* and *actIII* fragments were used to probe restriction digests of total DNA from a number of actinomycetes including known polyketide producers and some nonproducers. The results indicated that hybridisation bands were seen in nearly all of the producers and were absent from most of the nonproducers suggesting some level of correlation. These *act* genes and other methods used to isolate a number of type II PKS genes are summarised in table 1.4.

Host	**Polyketide**	**PKS genes**	**Cloning strategy**
S. coelicolor	actinorhodin	*act*	Complementation
S. rimosus	oxytetracycline	*otc*	Resistance / complementation
S. glaucescens	tetracenomycin	*tcm*	Complementation
S. violaceoruber	granaticin	*gra*	*act* probe
S. coelicolor	spore pigment	*whiE*	Complementation / *act* probe
S. peucetius	daunorubicin	*dps*	*act* / *tcm* probes
S. cinnamonensis	unknown	*mon*	*act* probe
S. halstedii	spore pigment	*sch*	*act* probe
S. curacoi	spore pigment	*cur*	*act* probe
Sac. hirsuta	unknown	*hir*	*act* probe
S. roseofulvus	frenolicin	*fren*	*act* probe
S. griseus	griseusin	*gris*	*act* probe
S. venezuelae	jadomycin	*jad*	*act* probe
S. spp. C5	daunorubicin	*dau*	*act* probe
Kib. aridum	unknown	*ard*	*act* probe
S. fradiae	urdamycin	*urd*	*tcm* / *act* probes
S. nogalater	nogalamycin	*sno*	*act* probe
S. argillaceus	mithramycin	*mtm*	*act* probe

Table 1.4. Cloning and sequencing of aromatic PKS gene clusters from actinomycetes (reproduced from Hopwood, 1997).

It was also found that spore pigmentation in *Streptomyces* spp. depends on the production of polycyclic aromatic polyketides during the maturation of the spores in the aerial mycelium. The first genetic locus involved in *Streptomyces* spp. spore pigmentation to be identified and cloned was the *whiE* cluster from *S. coelicolor* A3

(2) (Fu *et al.*, 1994). Mutations in the *whiE* gene cluster abolished or modified spore pigmentation. Some of the gene products resembled the protein subunits of PKSs already identified including those for actinorhodin (Tsoi and Khosla, 1995). Isolation and elucidation of the corresponding spore pigments has been unsuccessful, perhaps due to the fact that these metabolites polymerise and become covalently bound to macromolecular spore components (Yu *et al.*, 1998).

1.11. The minimal PKS

Initially regions from the granaticin (*gra*) and *tcm* PKS were sequenced (Bibb *et al.*, 1989; Sherman *et al.*, 1989) and they confirmed a type II nature of the PKS by showing open reading frames that would encode proteins resembling the condensation enzyme (the product of fabB) of the *E. coli* FAS and the discreet acyl carrier proteins of the type II FASs of bacteria and plants. Sequencing of the corresponding *actI* region revealed a similar genetic arrangement (Fernández-Moreno *et al.*, 1992). It has since been found that all identified gene clusters for actinomycetes encoding aromatic polyketides contain a set of three genes encoding the so-called "minimal PKS". This consists of a ketosynthase (KS or KSα), which also carries a putative acyltransferase (AT) domain, a chain length factor (CLF or KSβ) and an acyl carrier protein (ACP, McDaniel *et al.*, 1995). This minimal PKS is essential for polyketide biosynthesis and the lack of the KSα or KSβ (or ACP) results in an inactive PKS (Seow *et al.*, 1997). The arrangement of the *act* genes and the comparison to those that encode other aromatic PKSs is shown in Figure 1.6. There are similarities in the overall architecture between the various clusters,

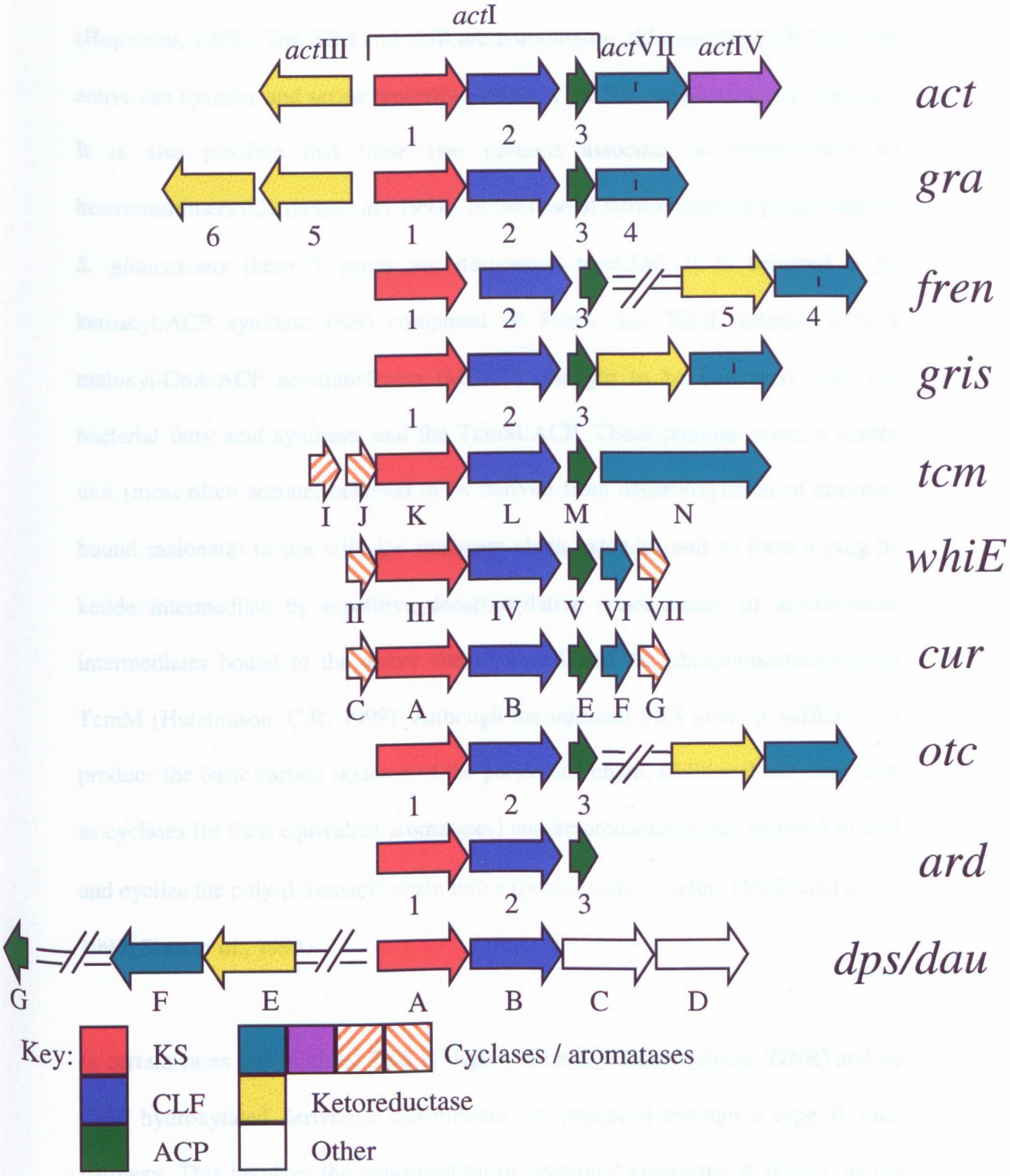


Figure 1.6. Organisation of gene clusters for type II PKSs in actinomycetes (Reproduced from Hopwood, 1997). For key see Table 1.4.

but also differences in the arrangements of some of the homologous genes (Hopwood, 1997). The KSα and KSβ are homologous although the KSβ lacks the active site cysteine and serine typically present in the KS and putative AT domains. It is also possible that these two proteins associate as heterodimers or heteromultimers (Carreras *et al.,* 1997). In the case of tetracenomycin production by *S. glaucescens* these 3 genes are designated *tcmKLM.* It is believed a β-ketoacyl:ACP synthase (KS) composed of TcmK and TcmL interact with a malonyl-CoA:ACP acyltransferase (MCAT), thought to be borrowed from the bacterial fatty acid synthase, and the TcmM:ACP. These proteins select a starter unit (most often acetate, believed to be derived from decarboxylation of enzyme-bound malonate) to use with the malonate chain extender unit to form a poly-β-ketide intermediate by repetitive decarboxylative condensation of acylthioester intermediates bound to the active site of TcmK and the phosphopantheinylayed TcmM (Hutchinson, C.R. 1999). Although the minimal PKS alone is sufficient to produce the basic carbon skeleton of the polyketide chain, additional enzymes such as cyclases (or their equivalent aromatases) and ketoreductases, are required to fold and cyclize the poly-β-ketoacyl chain into a (poly) cyclic structure (McDaniel *et al.*, 1995; Shen *et al.*, 1995).

In certain cases this minimal PKS is slightly different. Daunorubicin (DNR) and its C-14 hydroxylated derivative doxorubicin are produced through a type II PKS pathway. This involves the condensation of propionyl coenzyme A (CoA), as the starter unit, and nine malonyl-CoA extender units in the production of a 21-carbon decaketide (Grimm *et al.*, 1994). Compared with other type II PKS clusters the DNR PKS (*dps*) contains several unique features. Directly downstream of the genes

encoding the β-ketoacyl:acyl carrier protein synthase (*dpsA* and *dpsB*, KS) subunits are two unique genes, *dpsC* and *dpsD*, rather than an ACP which is found in all other type II PKS clusters. The ACP gene, *dpsG*, has an atypical position within the cluster, approximately 6.8 kb upstream of the genes encoding the KS subunits. It is believed that *dpsC* and *dpsD* play a crucial role in starter unit specificity (Bao *et al.*, 1999). Most KSα / KSβ gene pairs are translationally coupled except for two known exceptions. The homologous genes to *dpsA* and *dpsB* in *Streptomyces* sp. C5, a daunorubicin producer, are separated by approximately 100 bp and are not transcriptionally coupled (Ye *et al.*, 1994), as is the case in frenolicin production by *S. roseofulvus* (Bibb *et al.*, 1994). The KS gene pair, *dpsA* and *dpsB*, involved in daunorubicin production in *S. peucetius* are separated by only 3 bp so may be transcriptionally coupled (Grimm *et al.*, 1994).

1.12. Summary of polyketide research

1907 "The chemical synthesis of polyketides is centred around the reactive groups of the ketone and its adjacent α-methylene carbon" (Collie, 1907).

1945 Mechanism available for the biosynthesis of fatty acids including head-to-tail nomenclature for acetate condensations (Cornforth and Popják, 1945; quoted in Bentley and Bennett, 1999).

1953 Polyketones formed by head-to-tail linkage of acetate units followed by cyclization by aldol or acylation reactions (Birch and Donovan, 1953). The term polyketide was not used.

1957 The term macrolide (Makrolide) termed for recurring carbon patterns resulting from propionate units as shown by the structure of erythromycin (Woodward, 1957).

1960 First cell free extracts for the formation of a classical fungal polyketide, patulin from *Penicillium patulum* (formally *P. urticae*, Bassett and Tanenbaum, 1960).

1983 First gene for a secondary metabolite pathway cloned, an O-methyltransferase (Feitelson and Hopwood, 1983).

1984 *act* cluster cloned and expressed in *S. parvulus* (Malpartida and Hopwood, 1984).

1985 First hybrid antibiotics synthesised by moving actinorhodin genes form *S. coelicolor* into other *Streptomyces* species (Hopwood *et al.*, 1985).

1987 Tetracenomycin biosynthetic genes isolated from *S. glaucescens* (Motamedi and Hutchinson, 1987).

1987 DNA probes for *actI* and *actIII* found to hybridise to genomic DNA from 25 streptomyces and Southern analysis revealed bands for 14 out of 18 known polyketide producers (Malpartida *et al.*, 1987).

1988 Oxytetracycline biosynthetic genes isolated from *S. rimosus* (Butler *et al.*, 1989).

1989 Erythromycin PKS cloned by the Leadley group (University of Cambridge, Cambridge) using a self-resistance gene and the Katz group (Abbott Laboratories, Chicago) using mutant complementation (Cortes *et al.*, 1990: Donadio *et al.*, 1991).

1995 Design rules published for rational design of novel aromatic polyketides (McDaniel *et al.*, 1995).

1997 First example of a functional gene cloned and expressed from community DNA. (Seow *et al.*, 1997).

1999 Evidence for the "chain-length factor" or KSβ being a factor in polyketide chain initiation and starter unit specificity (Bisang *et al.*, 1999).

1.13. Aims

The overall aim of this study was to improve our current knowledge of the distribution and diversity of type II PKS genes in the natural environment, in particular the different soil types found in Cuba. Specifically;

- To develop molecular methods for the detection of type II PKS genes based on our current knowledge of the conservation of these gene clusters, and apply these methods to the natural environment.

- To study the distribution of type II PKS genes in total community DNA and in DNA extracted from culturable actinomycete isolates.

- To examine the phylogenetic diversity of actinomycete isolates containing type II PKS genes based on a number of genetic markers.

- To study the diversity of polyketide genes obtained directly from total community DNA to determine its usefulness in providing diverse sequences

- To development methods to extract novel type II PKS sequences from actinomycete isolates and total community DNA and exploit the potential of heterologous expression of these sequences to produce novel compounds.

Chapter 2

Materials and Methods

2.1. Bacterial strains

Actinomycete strains were stored as suspensions of spores and mycelial fragments in glycerol (10 % v / v) at -20°C (Wellington and Williams, 1978). *Escherichia coli* supercompetent cells were stored at -70°C in the manufacturers media (Invitrogen). Top 10F', DH5α and Dam⁻ competent cells were stored at -70°C in 50 µl aliquots after resuspension in ice cold glycerol (25 % v / v). See Table 2.1 for strain listing.

Table 2.1. Bacterial strains used

Bacterial Strains	Strain origin
Kibdelosporangium aridum	DSM 43828
Saccharopolyspora hirsuta	NCIMB 11079
S. cinnamonensis	NCIMB 12604
S. coelicolor A3(2)	Malpartida and Hopwood (1984)
S. curacoi	ATCC 19745
S. glaucescens	ETH 22794
S. griseus	ATCC 3478
S. halstedii	NRRL 2381
S. peucetius	NCIMB 9839
S. rimosus	NRRL 2234
S. roseofulvus	JCM 4334
S. violaceoruber	JCM 4979
Saccharopolyspora erythraea	ATCC 11635
NM 1-100	Cuban soil isolates
R 1-100	Cuban soil isolates
B 1-100	Cuban soil isolates
CB 1-50	Cuban soil isolates
Escherichia coli	Invitrogen TOP10F' cells

Escherichia coli	Invitrogen DH5α cells
Escherichia coli	Invitrogen Dam$^-$ cells

2.2. Vectors

The vectors and expression systems used and constructed in this study are listed in Table 2.2. Construction of the PKS expression vectors is discussed in Chapter 6.

Table 2.2. Vectors used

Vector	Origin
pCR 2.1 TA cloning vector	Invitrogen
pUC18 (*Sma*I cut)	Amersham Pharmacia
PKS expression vectors	Constructed in collaboration with P.F. Long and C.R. Hutchinson, University of Wisconsin (Madison)

2.3. Media

All media were made with distilled water unless otherwise stated and sterilised by autoclaving at 121°C for 15 min at 15 psi. Media and their constituents are listed in Table 2.3. Antibiotics used in this study are listed in Table 2.4.

Table 2.3. Media

Media	Constituent (s) l^{-1} (Unless otherwise stated)
IPTG stock	IPTG 238 mg 10 ml ddH_20 Stored in aliquots at -20°C
Luria Broth (LB)	Bacto-tryptone 10 g (DIFCO) Yeast extract 5 g (OXOID) NaCl 10 g (BDH) Adjusted to pH 7.0
LB agar	As above Agar 15 g (DIFCO)
Modified TSB	Tryptone soya broth 30 g (OXOID) Sucrose 100 g (Fisher Scientific)
Nutrient agar (NA)	Nutrient agar 28 g (OXOID)
Reduced Arginine Starch Salts (RASS) agar (Herron and Wellington, 1990)	L- arginine 0.1 g (Sigma) Soluble starch 12.5 g (BDH) K_2PO_3 1 g (BDH) NaCl 1 g (BDH) $MgSO_4.7H_2O$ 0.5 g (BDH) $FeSO_4.7H_2O$ (1% w / v) 1 ml (FISONS) $ZnSO_4.7H_20$ (1% w / v) 1 ml (FISONS) $CuSO_4.7H_2O$ (1% w / v) 1 ml (FISONS) $MgSO_4.7H_2O$ (1% w / v) 1 ml (BDH)

	Agar 15 g (DIFCO) Adjusted to pH 7.0
SOC	Tryptone 20 g (Sigma) Yeast extract 5 g (Sigma) NaCl 10.0 mM (Sigma) KCl 2.5 mM (Sigma) $MgCl_2.6H_2O$ 10.0 mM (Sigma) Glucose 20.0 mM (Sigma)
TAXI	1 LB (20 ml) agar plate (warmed to 37 °C for 10 min) X-Gal stock 40 µl IPTG stock 40 µl Ampicillin stock 20 µl Incubated at 37 °C for 30 min before use
Tryptone soya broth (TSB)	Tryptone soya broth 30 g (OXOID)
X-Gal stock	X-Gal 400 mg 10 ml dimethylformamide Stored at –20°C away from light

Table 2.4. Antibiotic stock solutions

Antibiotic	Stock solution (mg ml^{-1})	Final Concentration (µg ml^{-1})
Ampicillin[1]	50	50
Cycloheximide[2]	50	50
Nystatin[3]	50	50

Thiostrepton[4]	50	50

[1]Dissolved in SDW and filter sterilised (0.22 µm filter, Gelman Scientific).

[2]Dissolved in SDW and sterilised by autoclaving.

[3]Dissolved in a small volume of 0.1 M NaOH and made to volume with SDW.

[4]Dissolved in DMSO.

2.4. Buffers, reagents and solutions

Buffers, reagents and solutions were made with distilled water and sterilised by autoclaving at 121°C for 15 min at 15 psi. Buffers, reagents and solutions are listed in Table 2.5. Any additional buffers, reagents or solutions were prepared as described in Sambrook *et al.* (1989) and Hopwood *et al.* (1985).

Table 2.5. Buffers, reagents and solutions used in this study

Solution / Reagent	**Constituent (s) l^{-1}** (unless otherwise stated)
Alkaline solution	Sodium hydroxide 0.2 M Sodium dodecyl sulphate 1 % Made up fresh before use
Loading buffer	Sucrose 60% EDTA 100 mM Bromophenol blue 0.25% (w / v)

Lysis buffer	Tris 25 mM EDTA 25 mM Lysosyme 10 g Store at -20°C
Miniprep lysis buffer	Glucose 50 mM Tris-HCl 25 mM (pH 8) EDTA 10 mM Lysosyme 10 g RNase 60 ml
RNase stock solution	RNase 60 g heat treat at 100°C for 15 min store in aliquots at -20°C
¼ strength Ringers solution	NaCl 2.25 g KCl 0.105 g $CaCl_2$ 0.12 g $NaHCO_3$ 0.05 g
50 x TAE	Tris base 242 g Glacial acetic acid 0.5 M EDTA (pH 8.0) 100 ml
5 x TBE	Tris base 54 g Boric acid 27.5 g 0.5 M EDTA (pH 8.0) 20 ml
TE	Tris-HCL 10 mM EDTA 50 mM

2.5. Soil samples

Fresh soil samples were obtained from various sites across Cuba. Samples were placed into three groups, crop, botanical garden and Cayo Blanco. Rhizosphere samples were defined as soil attached to the plant roots. See Tables 3.1, 3.2 and 3.3 for a complete listing of soils studied.

2.6. Actinomycete and total viable isolations from soil

Air dried soil (1 g) was resuspended in 9 ml ¼ strength Ringers solution and shaken on a wrist action shaker for 10 min. A ten-fold dilution series was performed in ¼ strength Ringers solution and samples plated onto RASS (actinomycete counts) and NA (total culturable counts). Nystatin (50 $\mu g\ ml^{-1}$) and cycloheximide (50 $\mu g\ ml^{-1}$) were added to both media. Plates were incubated at 30°C for 7-14 d, and colonies enumerated. Presumptive actinomycete colonies were picked off, purified and cultured on RASS plates. Cultures were grown for 7-14 d and stored as spore suspensions (2.1).

2.7. Total community DNA isolation from soil

1 g of soil was suspended in 5 ml 0.12 M sodium phosphate buffer with 1 g Glaspergen glass beads (0.1-0.11 mm) in a Braun bead-beating bottle. Bead-beating was performed for 5 min with CO_2 cooling. The suspension was transferred to a sterile universal with 200 µl lysozyme (100 mg ml^{-1}) and placed in a shaking

incubator at 37°C. The suspension was pelleted by centrifugation for 15 min at 1660 x g and the supernatant transferred to a sterile universal. The pellet was re-extracted with 5 ml 0.12 M sodium phosphate buffer and after a further centrifugation for 15 min at 1600 x g the supernatants were combined. The sample was transferred to a sterile Oakridge tube, to which 5 ml of 8 M potassium acetate was added and the sample centrifuged at 13000 x g for 30 min. The supernatant was transferred to a sterile Oakridge tube, to which was added 10 ml PEG 6000 (50 % w / v) and 2 ml NaCl (5 M). The sample was incubated at 4°C overnight. The DNA was pelleted by centrifugation at 13800 x g for 30 minutes and the supernatant discarded. The pellet was resuspended in 1 ml TE and purified using Sephadex G50 (Sambrook *et al.*, 1989) and Chelex 100 columns (Straub *et al.*, 1994). This community DNA was found to be of sufficient purity to allow amplification of PCR products.

2.8. Standard DNA manipulations

2.8.1. Gel electrophoresis of DNA

Gel electrophoresis was used to visualise DNA fragments. Agarose gels were typically between 0.3 and 2.0% (w / v) containing 1 x TAE (or 1 x TBE) and 0.5 μg ml^{-1} ethidium bromide. Gels were visualised on a dual-intensity transilluminator (UVP Inc.). Photographs were taken with a Polaroid land camera and hood with a UV filter attachment using Polaroid 665 positive / negative film with a 5 sec exposure time.

2.8.2. DNA extraction from agarose gels

Gels were visualised on a dual-intensity transilluminator (UVP Inc.) and PCR products extracted from agarose gels using Qiagen gel extraction kits (Qiagen) or Geneclean II Kits (Bio101) as per manufacturers instructions. DNA fragments were also excised using electrophoresis onto DEAE-cellulose membrane and electroelution into dialysis bags as described in Sambrook *et al.* (1989).

2.9. Restriction endonuclease digestion of DNA

DNA was digested with restriction enzymes using the buffers and conditions as specified by the manufacturers (GibcoBRL and New England Biolabs). Digestions were performed in volumes between 10 and 50 µl and visualised as described in 2.8.1.

2.9.1. End repair of restriction products

Sticky ends provided by restriction digestion were repaired using T4 DNA polymerase. This either catalysed the synthesis of DNA in a 5'→3' direction or had 3'→5' exonuclease activity. This either allowed the removal of 3' overhangs or 5' overhangs to be filled in giving blunt ended products. This was achieved by incubating restriction products in the presence of 100 pM of each dNTP, 1-3 units T4 DNA polymerase and the relevant buffer (Gibco, BRL) at 12°C for 20 min as described in Sambrook *et al.* (1989).

2.10. Preparation of plasmid DNA

Small scale preparations of plasmid DNA were made using the QIAprep Spin Plasmid Kit (QIAGEN) using the buffers and conditions as specified by the manufacturers instructions.

2.10.1. Alkaline lysis plasmid preparations

Cells were grown overnight at 37°C in 3 ml LB with vigorous shaking. Cells were harvested from 2 ml of culture and spun down at 16060 x g for 2 min in a bench centrifuge. Cells were resuspended in 200 µl miniprep lysis buffer, vortexed and left at room temperature for 5 min. Alkaline solution (400 µl) was added and the solution mixed by inversion. 300 µl of 7.5 M ammonium acetate was added, the solution mixed by inversion and left on ice for 10 min. The solution was centrifuged at 16060 x g for 10 min and the supernatent added to 650 µl isopropanol. After incubation on ice for 15 min the DNA was collected by centrifugation at 16060 x g for 15 min. The pellet was washed in 70 % (v / v) ethanol and centrifuged at 16060 x for 5 min. The DNA pellet was dried in a vacuum desiccator and resuspended in 50 µl ddH_2O.

2.11. DNA quantification

Spectrophotometer readings were taken at 260nm and 280nm. An OD_{260} reading of 1.0 was taken to be equal to 50 mg ml^{-1} DNA. Purity was measured by the ratio OD_{260} / OD_{280}, a ratio of 1.8 was taken to be pure DNA (Sambrook *et al.*, 1989).

2.12. Small scale actinomycete DNA isolation

Actinomycete strains were inoculated into modified TSB (10 ml) in glass universal tubes. Cultures were incubated at 30°C for 5 d on an orbital shaker. Cells were harvested by centrifugation at 16060 x g for 10 min from 2 ml of culture, resuspended in 500 µl lysis buffer and incubated at 37°C for 2 h in an orbital shaker. SDS (125µl, 10 % w / v) was added and the samples incubated at 65°C for 30 min. 5 M potassium acetate (216 µl) was added and the samples mixed gently by inversion. Samples were placed on ice for a minimum of 30 min and centrifuged at 16060 x g for 10 min and the supernatant retained. This step was repeated to ensure no carryover of the pellet. An equal volume of isopropanol was added, the samples inverted 3-4 times and left at room temperature for 10 min. DNA was pelleted by centrifugation at 16060 x g for 20 min and the pellet washed in 75 % EtOH. DNA was resuspended in 50 µl TE and 1 µl RNase (60 mg ml^{-1}) and left overnight at 65°C to resuspend.

2.13. PCR amplification

All amplification reactions were carried out in 50 µl reaction volumes in 0.5 ml microfuge tubes. A standard PCR amplification protocol was followed for all PCR reactions, except reactions involving primers PFL1 and PFL2. Specific reaction conditions are shown in Table 2.6. Standard conditions consisted of 25 mM $MgCl_2$ (variable), DMSO 5.0 µl, BSA 5 µl (10 mg ml^{-1}), 10 x reaction buffer 5.0 µl, dNTPs 2 µl (10 mM of each dNTP), 100 ng template DNA, 100 ng of each primer and 2.5

units of *Taq* polymerase (Gibco BRL). Reaction volumes were made up to 50 µl using ddH_20. Reactions were carried out in Hybaid thermal cyclers under the following conditions: 1 cycle of 96°C for 5 min, 80 °C for 5 min ("Hot start"), annealing for 5 min and 72 °C for 2.5 min. This was followed by 35 cycles of 96 °C for 2 min, annealing for 2 min and 72 °C for 2.5 min. For PCR reactions involving PFL1 and PFL2 conditions consisted of 25 mM $MgCl_2$ (3 µl), DMSO (5.0 µl), 10 x reaction buffer (5.0 µl), dNTPs 2 µl (10 mM of each dNTP), 100 ng template DNA, 100 ng of each primer and 2.5 units of Vent polymerase (New England Biolabs) made up to 50 µl using ddH_20. PCR reactions were carried out in a Hybaid PCRExpress thermal cycler under the following conditions: 1 cycle of 95°C for 5 min, 80°C for 5 min, annealing for 5 min at 69°C and 72°C for 3 min. This was followed by a touchdown procedure of single cycles of 95°C for 1 min, annealing temperatures of 69-61°C for 2 min decreasing at 1°C per cycle followed by 72°C for 3 min. Final reaction conditions were 95 for 1 min, 61°C for 2 min and 72°C for 3 min for 27 cycles followed by a final extension time of 72°C for 10 min.

Table 2.6. Primers and conditions for PCR amplification

Primers	**Sequence**	**Target**	**Annealing / $MgCl_2$**	**Reference**
PKS1	5'-TTCGACGC CATCAAGGC CACC-3'	KSα	60°C/ 25 mM $MgCl_2$ (2.5 µl)	(This study)
PKS2	5'-GTCGGCGA AGACGACGTC	KSβ		

	GAC-3'			
SN1	5'-GACCCGGA GTGCGACCTG GAC-3'	KSα	60°C/ 25 mM $MgCl_2$ (2.5 µl)	(This study)
SN2	5'-GTCCAGGT CGCACTCCGG GTC-3'	KSβ		
pA	5'-AGAGT TTGATCC TGGCTC-3'	16S rDNA gene	58°C / 25 mM $MgCl_2$ (1.5 µl)	Edwards *et al.*, 1989
pH	5'-AAGGA GGTGATCC AGCCGCA-3'			
A1	5'-GGATGAG CCCGCGGCC TA-3'	16S rDNA gene	58°C 25 mM $MgCl_2$ (1.5 µl)	Heuer *et al.*, 1997
A2	5'-GCCGGCC GCGGCTGCTG GCACGTA-3'			
RecA1	5'-GT(CG)GA GATCTACGG CCC(CG)GA-3'	*recA*	52°C / 25 mM $MgCl_2$ (1.5 µl)	Nuβbaumer & Wohlleben, 1994
RecA2	5'-GCGTAGA ACTT(CG)AG (CG)GCGTTGC CGCC-3'			
PFL1	5'-C<u>CTGCAG</u>[1]G GAGG**ATG**[2]CG	KSα	6.2	University of

	NCGSGTCGTCA TCACCGGCATC GG-3'			Wisconsin, Madison
PFL2	5'-CTCCGGA[3]**TC** **A**[4]CGCSGAGTTG AASCCGCCGTG GCCGCGSGC-3'	KSβ	6.2	
M13 (F)	5'-GTAAAACGAC GGCCAGT-3'	Cloning Vectors	2.15	Invitrogen
M13 (R)	5'-AACAGCTATG ACCATG-3'			Invitrogen
T7	5'-AATACGACTC ACTATAG-3'	Cloning Vectors	2.15	Invitrogen

1 - CTG CAG, *Pst*I site
2 – **ATG**, transcriptional start site
3 – TCC GGA, *Mro*I site
4 - **TCA**, stop site (TGA)

2.13.1. Phosphorylation of PCR products

PCR products amplified using Vent polymerase (New England Biolabs) were phosphorylated after gel extraction. Extracted products were made up to 103 µl using ddH_2O and added to 30 µl 5 x reaction buffer (Gibco BRL), 15 µl ATP (100 mM) and 2 µl Kinase (Gibco BRL) to give a final reaction volume of 150 µl. Phosphorylation was carried out at 37°C for 1 hour.

2.13.2. Phenol chloroform extraction

Phosphorylated products were cleaned up and precipitated using phenol chloroform precipitation. An equal volume of water saturated phenol was added and the samples agitated by vortexing for 2 min and centrifugation at 16060 x g for 2 min. Phenol was extracted from the solution by adding an equal volume of chloroform-isoamyl alcohol (24:1) to the removed upper aqueous phase. The samples were agitated by vortexing for 2 minutes and centrifugation for 2 min at 16060 x g. The upper aqueous phases were added to 0.1 volume sodium acetate (7.5 M) and 2.2 volume ethanol (100 %). Samples were vortexed and incubated at –20°C for 15 minutes followed by centrifugation at 16060 x g for 15 min. The pellets were washed with ethanol (70 %) and centrifuged at 16060 x g 5 min. Pellets were aspirated and dried by vacuum desiccation. Samples were resuspended in 30 µl of ddH_2O.

2.14. Cloning

PCR products amplified using Taq polymerase (Gibco BRL) were cloned using the Invitrogen TA cloning kit using the pCR 2.1 cloning vector and Top 10F' competent cells (Invitrogen) as per the manufacturers instructions. PCR products amplified using Vent polymerase (New England Biolabs) and other DNA fragments were cloned as outlined in 2.14.1 and 2.14.2.

2.14.1. Ligation reactions

Ligation reactions were set up in a total reaction volume of 20 µl comprising of 4 µl 5 x T4 ligase buffer (Gibco BRL), 1 µl vector (Amersham Pharmacia), 14 µl DNA and 1 µl T4 ligase (Gibco BRL). The vector used was dephosphorylated *Sma*I cut pUC18 (Amersham Pharmacia). Vector and DNA concentrations were taken from Sambrook *et al.* (1989). Ligation reactions were incubated at 14 °C overnight.

2.14.2. Transformation reactions

Transformations were carried out using competent *E. coli* DH5α and Dam$^-$ cells (Invitrogen) prepared as described in 2.14.3. Using a chilled sterile pipette ligation mixes were added to cell aliquots (50 µl), mixed gently and stored on ice for 30 min. Cells were heat shocked for 90 sec at 42°C and left on ice for 2 min. LB (500 µl) was added and samples incubated at 37°C for 45 min with shaking. Aliquots were then plated on to LB agar containing the relevant antibiotic, X-Gal and IPTG (TAXI plates). TAXI plates were incubated overnight at 37°C. Antibiotic, X-Gal and IPTG concentrations were taken from Sambrook *et al.* (1989).

2.14.3. Preparation of competent cells

Competent cells were prepared from *E. coli* DH5α, Dam$^-$ and Top10F' stocks (Invitrogen). Cells were streaked onto LB agar plates and incubated at 37°C overnight. Two colonies were picked and dispersed in 1 ml SOB media (Sambrook *et al.*, 1989). This was used to inoculate 100 ml SOB which was incubated at 37°C

with shaking until an optical density of 0.6 was reached. The culture was transferred to ice cold Falcon tubes and left on ice for 10 min. The culture was centrifuged at 130 x g for 10 min (4 °C) and the supernatant decanted. The pellet was resuspended in 10 ml ice cold 0.1 M $CaCl_2$, vortexed and left on ice for 30 min. This was repeated twice with final resuspention in 2 ml volume of ice cold 0.1 M $CaCl_2$. Cells were incubated on ice for 3 hours and used directly. For storage an equal volume of 50 % glycerol was added and 50 µl aliquots stored at -70°C.

2.15. DNA sequencing and analysis

DNA sequencing reactions were carried out by cycle sequencing with the Dye Terminator Kit of PE Applied Biosystems (Warrington, Cheshire, UK). Total reaction volume (20 µl) included 2 µl DMSO, 2 µl reaction mix, 5 µl primer (1.5 pM μl^{-1}), 2.5 µl DNA and made up to total volume using ddH_20 and reaction buffer as per the manufacturers instructions (PE Applied Biosystems). Sequencing was carried out using T7 and M13 (forward and reverse) sequencing primers as published by Invitrogen and certain PCR primers as required. The sequencing thermal cycling program used was as recommended by PE Applied Biosystems. Sequencing reactions were purified using AutoSeq G50 columns (Amersham Pharmacia) and dried using a vacuum desiccator. Nucleotide sequences were analysed in BLAST 1.0 or 2.0 (http:// www.ncbi.nlm.nih.gov/cgi-bin/BLAST/) and aligned in ClustalW (http://www2.ebi.ac.uk./clustalW/).

2.16. Phylogenetic analysis

Nucleotide and peptide sequences were aligned using ClustalW (http://www2.ebi.ac.uk./clustalW/), and dendrograms constructed using the programs Dnadist, Seqboot, Neighbor and Consence from the PHYLIP software package (version 3.5, Felsenstein, 1985). Trees were prepared using Treeview (version 3.5, //taxonomy.zoology.gla.ac.uk/rod/treeview.html). Bootstrap values were indicated by the method of Felsenstein (1985) as implicated in PHYLIP.

Chapter 3

The development of a molecular method for the detection of type II polyketide synthase genes (PKSs) in soil community DNA

3.1. Introduction

Recombinant DNA and molecular phylogenetic methods have recently provided means for identifying the types of organisms that occur in microbial communities without the need for cultivation (Amann *et al.*, 1995). Results from the application of these molecular methods to a number of diverse environments confirm that our view of microbial diversity was limited and points to a wealth of novel and environmentally important diversity yet to be studied (Pace, 1997). It is estimated that >99 % of microorganisms observable in nature typically are not culturable using standard techniques (Amann *et al.*, 1995). Our knowledge of microbial diversity has increased dramatically in recent years, in part as a result of sequencing rRNA genes from DNA directly from uncultured microbiota. This approach has been applied to assess the microbial diversity in a variety of environments (Tanner *et al.*, 1998). The use of molecular ecology has allowed the following areas to be addressed (Stackebrandt, 1992);

- Identifying and analysing the natural relationships of organisms in mixed microbial populations.
- Studying the change of genetic complexity of microbial communities under changing environmental conditions (e.g. changes in nutrition and physiochemical parameters).
- Detecting organisms with specific metabolic activities such as N_2 fixation, bioremediation, secondary metabolite production, waste degradation, heavy metal accumulation and biotransformation.

- Monitoring the exchange of genetic material between species (horizontal gene transfer), usually only possible when studying isolates.
- Monitoring the maintenance of marker genes in genetically engineered microorganisms (GEMs) in the natural environment.

The major strategies for characterising microbial communities without the need for cultivation are outlined in Figure 3.1, adapted from Hugenholtz and Pace (1996). The development of the polymerase chain reaction (PCR, Mullis and Faloona, 1987) was a major methodological breakthrough enhancing the probability of detecting rare sequences in heterologous mixtures of DNA which previous hybridisation studies had not allowed (Steffan and Atlas, 1988).

Soil represents probably one of the most complex and difficult environments to study. There is evidence indicating there might be several thousand microbial species in one soil sample (Torsvik *et al.*, 1990). Recent analysis of 16S rDNA clone libraries generated from DNA isolated from terrestrial and marine environments indicated the existence of a broad monophyletic group of hitherto uncultured organisms falling within the actinomycete line of descent (Rheims *et al.*, 1996). Due to the use of PCR to amplify small amounts of DNA, organisms occurring in small numbers in the environment are now detectable (Wintzingerode *et al.*, 1997). The sample volume required for analysis is significantly reduced and micro-habitats are now open for investigation e.g. termite guts (Ohkuma *et al.*, 1996) and nitrogen-fixing root nodules (Mirza *et al.*, 1994).

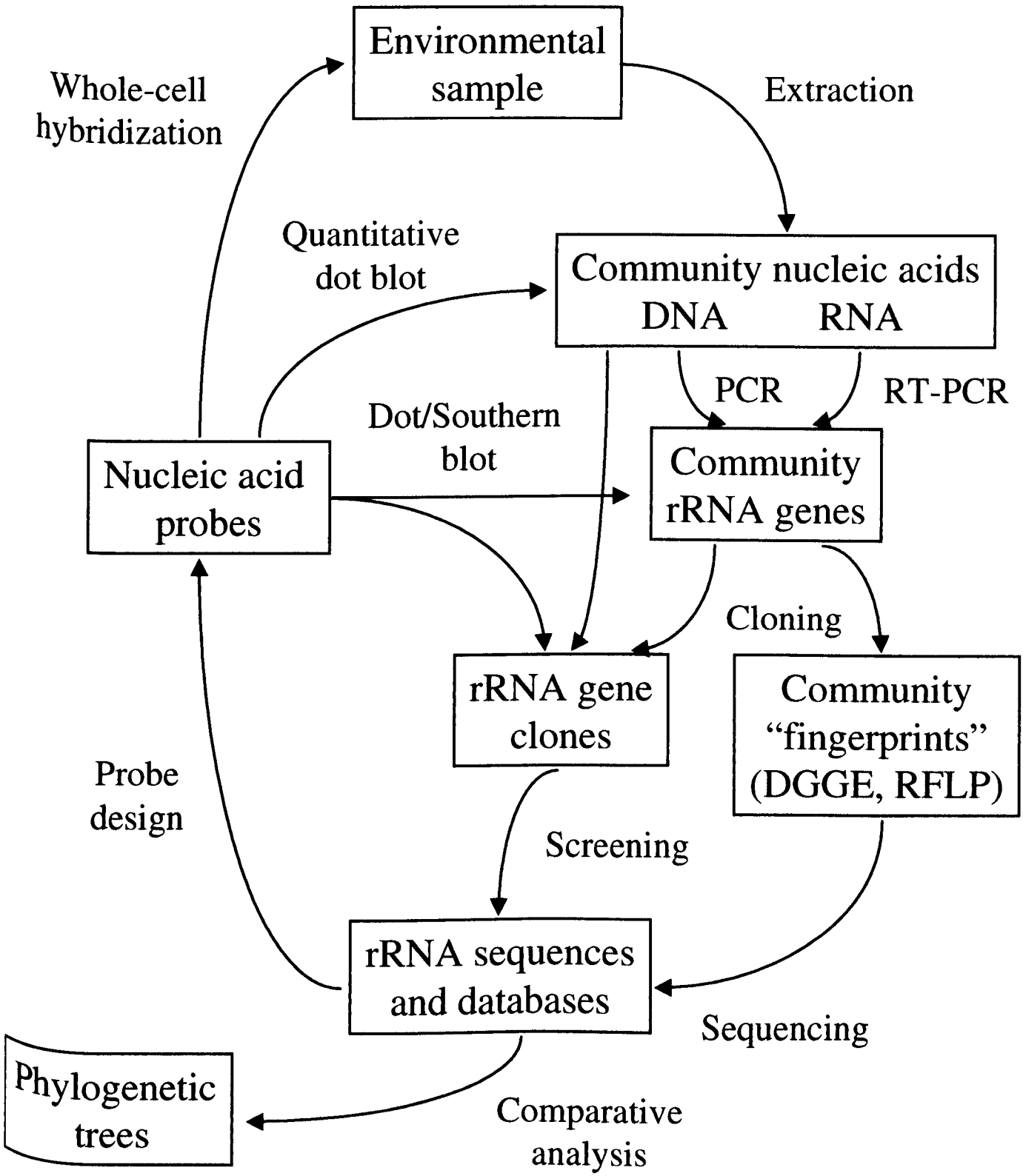


Figure 3.1. Strategies for characterizing microbial communities without cultivation (Reproduced from Hugenholtz and Pace, 1996).

Isolation of total community DNA is the first step for studying molecular ecology. Direct extraction of DNA from soil yields large amounts of DNA but it may contain impurities that can inhibit subsequent enzymatic reactions (Steffan *et al.*, 1998). Because the yield is so great, however, more purification steps may be incorporated with a lower percentage of DNA loss (Steffen and Atlas, 1988). Lysis of microbial cells from environmental habitats is a critical step. Insufficient or preferential disruption of cells will bias the view of the microbial diversity. Rigorous conditions required for cell lysis of Gram positive bacteria should be avoided as the treatment may give rise to fragmented nucleic acids leading to the formation of PCR generated artefacts or chimeric products (Wintzingerode *et al.*, 1997). A number of studies have addressed the use of different methods for the extraction of total community DNA from soil samples. These include a bead beating method followed by subsequent incubation in hot (70 °C) sodium dodecyl sulphate (Ogram *et al.*, 1987), lysozyme treatment and rapid freeze thawing (Tsai and Olsen, 1991) and a cation exchange indirect lysis approach (Jacobsen and Rasmussen, 1992). The most efficient and reliable method for the lysis of actinomycete spores and mycelia is mechanical lysis by bead beating (Morris *et al.*, 1998).

There are a number of methods for the purification of total community DNA once it has been extracted. Methods include ion-exchange resins such as Chelex 100 (Straub *et al.*, 1994), Sephadex G50 and CsCl-ethidium bromide density gradient ultracentrification (Holben *et al.*, 1988), phenol chloroform extraction (Fuhrman *et al.*, 1988) and polyvinylpolypyrrolidone (PVPP) treatments (Barns *et al.*, 1994). These purification methods may help to remove humic acids and other humic

substances which may inhibit *Taq* polymerase (Tebbe *et al.*, 1993). These substances may also interfere with other DNA modifying enzymes such as restriction enzymes and DNaseI (Porteous *et al.*, 1991). These substances may also interfere with DNA hybridridisation specificities (Steffan and Atlas, 1988). In contrast to terrestrial habitats aquatic environments contain significantly lower levels of inorganic or organic particles and lysis protocols developed for pure cultures have been successfully applied (Somerville *et al.*, 1989).

Molecular tools have great potential to assist in the identification of uncultured bacteria and to explore microbial diversity (Teske *et al.*, 1996). One of the molecular tools used is the PCR amplification of the genes encoding 16S rDNA by the use of primers homologous to conserved regions of the gene (Heuer *et al.*, 1997). The 23S rRNA gene is too large to be sequenced routinely and there are limits to the phylogenetic usefulness of the 5S rRNA gene (Stackebrandt, 1992). These ribosomal genes provide good markers for determining microbial diversity but there are certain drawbacks to there use. These include presence of contaminating DNA, 16S rRNA sequence variations due to rRNA operon heterogeneity and formation of deletion mutants due to excessive secondary structure (Wintzingerode *et al.*, 1997). Other molecular markers can be used to study specific genes. Examples include detection of methanol dehydrogenase genes (*mxaF*, McDonald and Murrell, 1997), naphthalene degrading genes (Herrick *et al.*, 1993), nitrogen reductase genes (*nifH*, Kirshtein *et al.*, 1991) and polychlorinated biphenyl-degrading catabolic genes (Walia *et al.*, 1990). Antibiotic biosynthesis and resistance genes have also been used as molecular markers in recent studies (Huddleston *et al.*, 1997; Seow *et al.*, 1997; Standage, 1998; and Wiener *et al.*,

1998). Huddleston *et al.* (1997) showed that total community DNA contained the aminoglycoside phosphotransferase gene, *strA*, for streptomycin resistance. Standage (1998) studied the presence of the gentamicin resistance methyltransferase gene (*grmA*) in a wide range of total community DNA samples.

The genetic markers chosen for primer design in this study were the α and β β-ketoacylsynthase (KS) genes present in all known type II PKSs. Previous studies have used degenerate oligonucleotide primers designed to amplify a part of the KSβ from a highly conserved region. These primers were used to screen *Streptomyces* spp. and bacterial strains enriched from soil samples. Using this primer set PCR products were found to show a distribution of spore and antibiotic PKSs which grouped phylogenetically into different classes. Antibiotic PKSs were grouped together on the basis of starter unit incorporation (Metsä-Ketelä *et al.*, 1999). A similar approach was undertaken by Seow *et al.* (1997) using PCR primers designed to amplify the whole KSβ open reading frame. These primers were used to amplify KSβ genes from unclassified actinomycete isolates and total community DNA. Sequence analysis of these PCR products showed them to be phylogenetically distinct when compared to known KSβ genes from both antibiotic and spore pigment type II PKSs.

3.1.1. Aims

- To develop molecular detection methods for type II polyketide synthase α and β β-ketoacylsynthase genes in natural communities.

- To study the distribution of type II polyketide synthase genes in terrestrial microbial communities using molecular techniques.

3.2. Development of molecular methods for the detection of type II polyketide synthase genes (PKSs)

A PCR based detection method for the detection of type II PKS genes was adapted from a previous study (K. Maycroft and W.P. Revill, pers. comm.). PCR primers, PKS 1 and PKS 2, were designed from an amino acid pileup sequence from 12 type II PKS sequences. A similarity plot of these 12 sequences allowed consensus primers to be designed to a region of approximately 1.6 Kb spanning the α and β β-ketoacylsynthase genes (Figure 3.2, Table 3.1).

Strain	type II PKS product	PCR product size
Kibdelosporangium aridum	unknown function	1550 bp
Saccharopolyspora hirsuta	unknown function	1529 bp
S. cinnamonensis	unknown function	1529 bp
S. coelicolor A3(2)	actinorhodin PKS	1541 bp
S. coelicolor A3(2)	*whiE* spore pigment	1577 bp
S. curacoi	probable spore pigment	1574 bp
S. glaucescens	tetracenomycin PKS	1550 bp
S. griseus	griseusin PKS	1562 bp
S. halstedii	spore pigment	1550 bp
S. rimosus	oxytetracycline PKS	1586 bp
S. roseofulvus	frenolycin PKS	1679 bp
S. violaceoruber	granaticin PKS	1544 bp

Table 3.1. The 12 type II PKS genes used for the construction of the PCR primers. The pileup sequence consists of 9 *Streptomyces* spp. and 2 other actinomycetes. *S. coelicolor* is represented twice, once for the *whiE* locus and once for the actinorhodin cluster.

A)

Kibdelosporangium aridum	5'-TTCGACGCGATCAAGGCGACC-3'
Saccharopolyspora hirsuta	5'-TTCGACGCCATCAAGGCGACC-3'
S. cinnamonensis	5'-CTCGACGCGATCAAGGCGACC-3'
S. coelicolor actI	5'-TTCGACGCGATCCGCGCCACG-3'
S. coelicolor whiE	5'-TTCGACGCCATCAAGGCGACC-3'
S. curacoi	5'-TTCGACGCCATCAAGGCCACG-3'
S. glaucescens	5'-TTCGACGCCATCAAGGCGACG-3'
S. griseus	5'-TTCGACGCGATCCGCGCCACC-3'
S. halstedii	5'-TTCGACGCCATCAAGGCGACC-3'
S. rimosus	5'-TTCGACGCGATCAAGGCCACC-3'
S. roseofulvus	5'-TTCGACGCCATCAAGGCCACC-3'
S. violaceoruber	5'-TTCGACGCCATCAAGGCGACG-3'
PKS 1	5'-TTCGACGCCATCAAGGCCACC-3'

B)

Kibdelosporangium aridum	5'-GTCGGCGAAAACGACGTCGAT-3'
Saccharopolyspora hirsuta	5'-GTCGGCGAACACCACGTCGAC-3'
S. cinnamonensis	5'-GTCGGCGAAGACACGTCGGAT-3'
S. coelicolor actI	5'-GTCGGCGAACACGACGTCGAC-3'
S. coelicolor whiE	5'-GTCCGCGAAGACCACGTCGAC-3'
S. curacoi	5'-GTCGGCGAAGACCACGTCGAT-3'
S. glaucescens	5'-GTCGGCGTAGACCACGTCGAC-3'
S. griseus	5'-GTCGGCGAAGACGACGTCCAC-3'
S. halstedii	5'-GTCGGCGAACACGACGTCCAC-3'
S. rimosus	5'-GTCGGCGAAGACCACGTCCAC-3'
S. roseofulvus	5'-GTCCGCGAAGACGACGTCGAC-3'
S. violaceoruber	5'-GTCGGCGAACACCACCGAGAT-3'
PKS 2	5'-GTCGGCGAAGACGACGTCGAC-3'

Figure 3.2. Pileup sequences determining the design of consensus PCR primers for amplification of type II PKS genes. A) Forward design, B) Reverse design. Letters in red denote conserved sequences with unconserved nucleotides represented in green and blue.

Forward and reverse semi-nested primers (SN1 and SN2) were designed from the same pileup sequence for product conformation. Figure 3.3. illustrates the primers and amplification of type II PKS sequences from actinomycete type strains.

3.3. Preparation of community DNA for molecular screening

Total community DNA was extracted from 65 soils using a method consisting primarily of beadbeating and lysozyme steps for spore and cell lysis (Herron and Wellington, 1990) and Sephadex G50 and Chelex 100 columns (Straub, 1994) for DNA purification (2.7). Community DNA was tested for purity by the ability to amplify 16S rRNA genes. Amplification was carried out using universal bacterial primers , pA and pH (Edwards *et al.*, 1989), and high G + C specific primers, A1 and A2 (Heuer *et al.*, 1997). The 16S rRNA gene could be amplified from all 65 soils tested using both primer sets. An example of this is shown in Figure 3.4.

3.4. Screening community DNA for type II PKS genes

Fresh soil samples (65) were taken from various sites around Cuba. Soil was frozen on arrival in order to keep it fresh. Sampling sites were placed into 3 catagories: crop soils, samples from a botanical garden site in Havana, and samples from Cayo Blanco. Cayo Blanco is an uninhabited pristine island located off the Cuban coast.

Polyketide (PKS 1 and PKS 2) specific primers were used to screen total community DNA for the presence of α and β KS genes. Semi-nested PCR using SN1 and SN2 was used to confirm PKS PCR products where a product of the

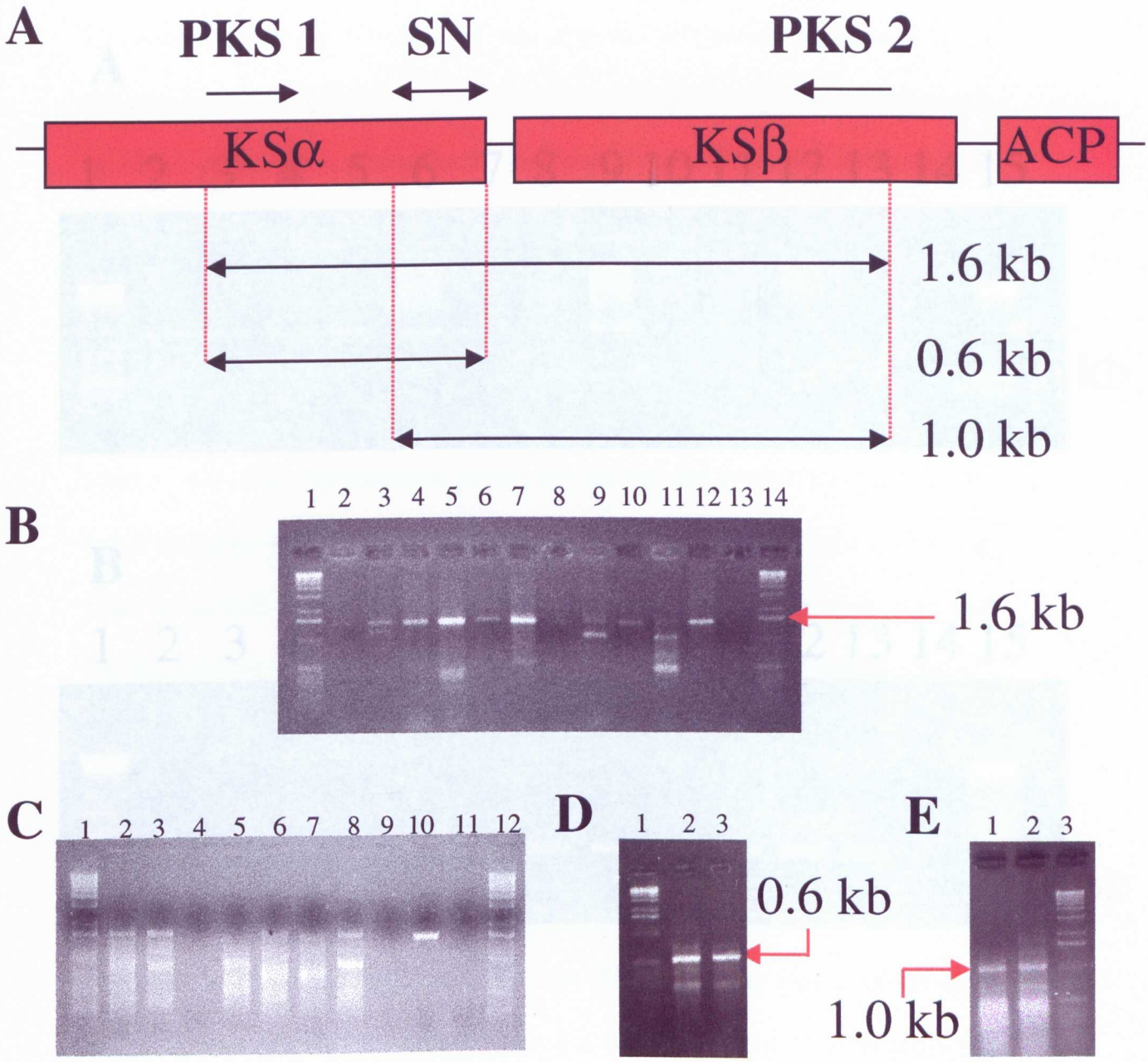


Figure 3.3. (A) Primer sets for the amplification of the α and β β-ketoacylsynthases from the type II minimal PKS. (B) Amplification of the α and β β-ketoacylsynthase genes using primers PKS 1 and PKS 2. Lane 1. 1 kb marker, 2. *Kibdelosporangium aridum*, 3. *Saccharopolyspora hirsuta*, 4. *S. cinnamonensis*, 5. *S. coelicolor* A3(2), 6. *S. curacoi*, 7. *S. glaucescens*, 8. *S. griseus*, 9. *S. halstedii*, 10. *S. rimosus*, 11. *S. roseofulvus*, 12. *S. violaceoruber*, 13. ddH_20, 14. 1 kb marker. **(C) Amplification of the α and β β-ketoacylsynthase genes using primers PKS 1 and PKS 2 from total community DNA.** Lane 1. 1 kb marker, 2-9. Total community DNA samples, 10. *S. coelicolor*, 11. ddH_20, 12. 1 kb marker. Positive bands observed in lanes 1,2,6 and 8. **(D) Semi nested PCR using PKS 1 and SN1 to confirm PKS PCR products amplified from total community DNA.** Lane 1. 1 kb marker, 2-3. Total community DNA PCR products. **(E) Semi nested PCR using PKS 1 and SN2 to confirm PKS PCR products amplified from total community DNA. Lane 1-2.** Total community DNA PCR products, 3. 1 kb marker.

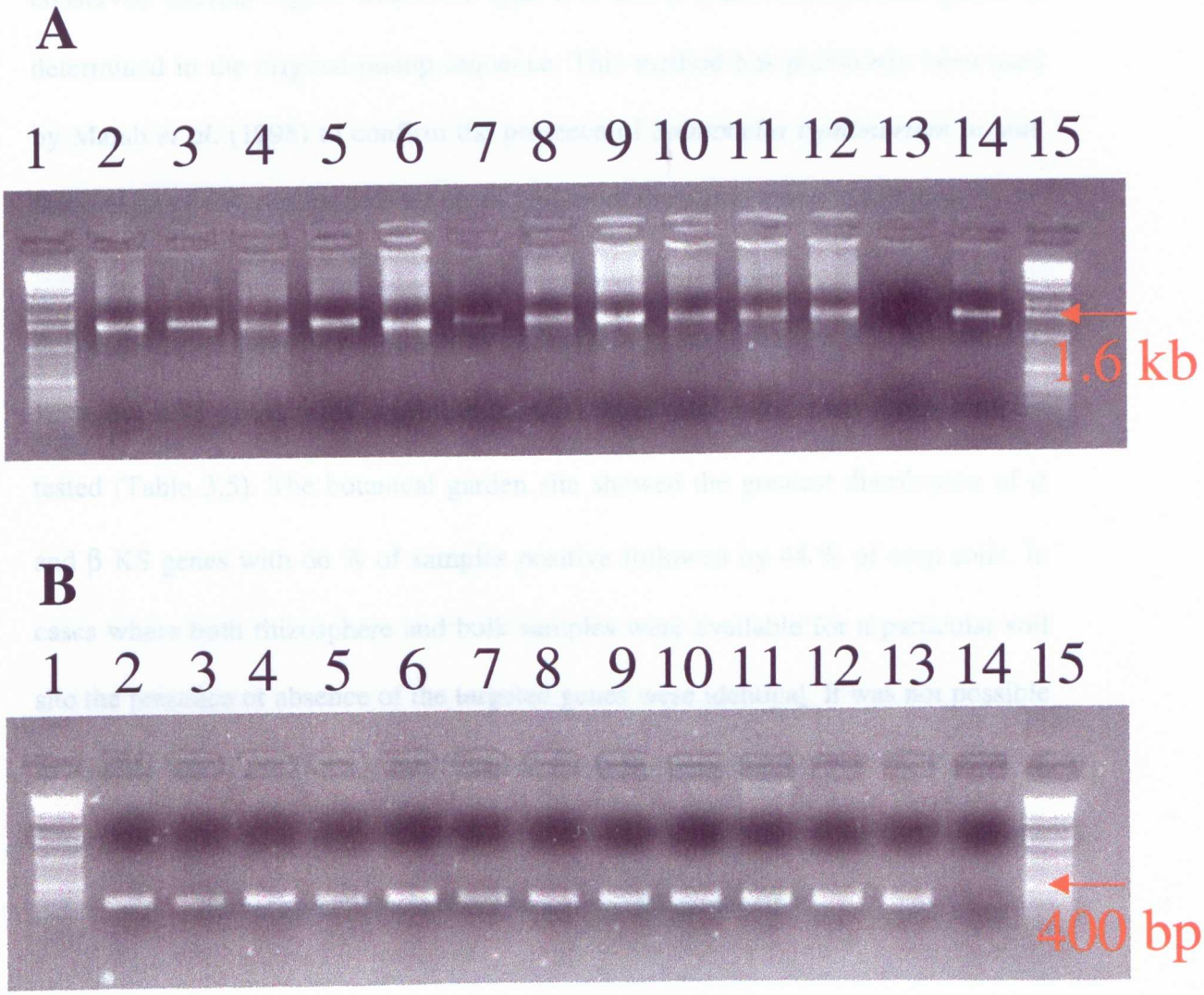


Figure 3.4. (A) Amplification of 16S rDNA genes from total community DNA using universal oligonucleotide primers (pA and pH, Edwards et al., 1989). Lane 1. 1 kb marker, 2-12. total community DNA samples, 13. ddH_2O, 14. *S. coelicolor*, 15. 1 kb marker. **(B) Amplification of 16S rDNA genes from total community DNA using high G + C specific primers (A1 and A2, Heuer et al., 1997).** Lane 1. 1 kb marker, 2-12. total community DNA samples, 13. *S. coelicolor*, 14. ddH_2O, 15. 1 kb ladder.

correct size was observed (~1.6 kb). These semi-nested primers were designed to a conserved internal region within the type II α and β β-ketoacylsynthase genes as determined in the original pileup sequence. This method has previously been used by Marsh *et al.* (1998) to confirm the presence of *Salmonella typhimurium* in soil. Semi-nested PCR was carried out on an aliquot of the initial reaction mixture.

A full listing of the soil communities screened is given in Tables 3.2, 3.3 and 3.4. KSα and KSβ genes were amplified from 43 % of total community DNA samples tested (Table 3.5). The botanical garden site showed the greatest distribution of α and β KS genes with 66 % of samples positive followed by 44 % of crop soils. In cases where both rhizosphere and bulk samples were available for a particular soil site the presence or absence of the targeted genes were identical. It was not possible to amplify any α and β KS genes from the Cayo Blanco samples.

3.5. Actinomycete isolation from Cayo Blanco soils

Selective isolation was applied to 8 Cayo Blanco soil samples. Actinomycete isolations were carried out on RASS agar following ¼ strength Ringers extraction (2.6). 50 actinomyctes were picked, grown up in TSB and subjected to a small scale DNA extraction procedure. This isolate DNA was tested for the presence of PKS genes using PCR with PKS specific primers (PKS 1 and PKS 2). Ninety percent of Cayo Blanco isolates were found to contain α and β KS genes.

3.6. Community DNA purification

Total community DNA samples from Cayo Blanco sites were further purified by agarose gel electrophoresis and subsequent gel extraction. DNA samples were adjusted to concentration of 100 ng μl^{-1} concentrations and subjected to PCR using the polyketide specific primers (PKS 1 and PKS 2). It was not possible to amplify α and β KS genes from these purified Cayo Blanco total community DNA samples.

3.7. PCR inhibition

PCR was carried out on *S. coelicor* DNA (100 ng μl^{-1}) using PKS specific primers (PKS 1 and PKS 2) and a dilution series of *S. coelicolor* DNA using either ddH_2O or Cayo Blanco total community DNA. *S. coelicolor* DNA samples were diluted from 100 ng μl^{-1} to 0.1 pg μl^{-1} using either ddH_2O or Cayo Blanco total community DNA and 1 µl used of each dilution per PCR reaction. These samples were subjected to PCR using the PCR specific PKS primers. In samples where *S. coelicolor* DNA was diluted using ddH_2O the detection limit was found to be 10 pg of total DNA (10 pg μl^{-1} dilution). In samples diluted using Cayo Blanco total community DNA only the control lane gave the correct amplification product (Figure 3.5). This experiment was repeated using *S. cinnamonensis* DNA and total community DNA from a different Cayo Blanco site and the same result was observed.

Soil type and crop (reference number)	PKS PCR amplification and product confirmation	Soil type and crop (reference number)	PKS PCR amplification and product confirmation
Black Bean (1)	-	Sugar cane (B, 202)	✓
Sweet potato (3)	-	Papaya (R, 203)	-
Red pepper (4)	✓	Papaya (B, 204)	-
Sugar cane (5)	-	Bean (R, 205)	✓
Havana (6)	✓	Bean (B, 206)	✓
Henequen (7)	-	Maize (R, 207)	-
Palm (8)	✓	Maize (B, 208)	-
Sugar cane (R, 201)	✓	Orange grove (209)	-

Table 3.2. Screening of crop soil total community DNA samples by PCR amplification and subsequent product confirmation. Where samples were available both rhizosphere (R) and bulk (B) soil were tested.

Plant type (reference number)	PKS PCR amplification and product confirmation	Plant type (reference number)	PKS PCR amplification and product confirmation
Botanical site	✓	Moa (227)	-
Franboyan (210)	✓	Marmelo (228)	✓
Pino (211)	✓	Café (229)	✓
Isora (212)	✓	Cacoa (230)	✓
Jaquae (213)	✓	Guayaba (231)	✓
Almasego (214)	✓	Manderina (233)	-
Bambi (215)	✓	Masaqua (234)	✓
Zico (216)	✓	Moa (235)	-
Yagramas (217)	✓	Guasima (236)	✓
Jaquey (218)	✓	Guasima (237)	-
Palmas (219)	✓	Caoba (238)	-
Mango (220)	-	Café (239)	-
Mango (221)	-	Mayagua (240)	✓
Caimyto (222)	-	Cocoa (241)	-
Guayabo (223)	✓	Caoba (242)	-
Haemelo (224)	✓		
Manderena (225)	✓		

Table 3.3. Screening of Botanical garden soil total community DNA samples by PCR amplification and subsequent product confirmation.

Soil type and crop (reference number)	PKS PCR amplification and product confirmation
Fir, rhizosphere (243)	-
Fir, rhizosphere (244)	-
Fir, rhizosphere (245)	-
Fir, rhizosphere (246)	-
Fir, rhizosphere (247)	-
Fir, rhizosphere (248)	-
Fir, rhizosphere (249)	-
Shrub, rhizosphere (250)	-
Shrub, away from rhizosphere (251)	-
Shrub, away from rhizosphere(252)	-
Shrub, away from rhizosphere(253)	-
Shrub, rhizosphere (254)	-
Shrub, rhizosphere (255)	-
Cayo Blanco (A)	-
Cayo Blanco (B)	-
Cayo Blanco (C)	-
Cayo Blanco (D)	-

Table 3.4. Screening of Cayo Blanco soil total community DNA samples by PCR amplification and subsequent product confirmation.

Soil type	% positive
Total	43 %
Botanical garden	66 %
Cayo Blanco	0 %
Crop soils	44 %

Table 3.5. Results from screening total community DNA samples by PCR analysis using type II PKS specific oligonucleotide primers. Soils were grouped into 3 types and the results given in terms of percentage positives.

3.8. Discussion

A PCR based detection method was developed to amplify a region of 1.6 kb spanning the KSα and KSβ regions of type II PKS gene clusters. This primer pair was designed from an amino acid pileup sequence of 12 type II PKS sequence and used to successfully amplify the 2 KS genes from the 11 type strains used in the primer design. An internal primer was designed from the same 12 sequences to be used as a semi-nested primer for product confirmation.

Total community DNA was extracted from a number of soil samples representative of different sites across Cuba. These included crop soils, samples from a Botanical garden located in Havana and a pristine island site, Cayo Blanco. KSα and KSβ genes could be amplified from 43 % of total community DNA samples. These genes were predominately amplified from Botanical garden sites with 66 % of these total community DNA samples giving amplification of KSα and KSβ gene pairs. It was possible to amplify KSα and KSβ genes from 44 % of total community DNA samples extracted from Cuban crop soils. In cases where bulk and rhizosphere samples were available for a particular site no difference was observed in the presence or absence of the targeted genes. This suggests a similar community present in both rhizosphere and bulk soils at levels above the detection limits for the PCR protocol.

In subsequent studies (Chapter 4) with actinomycete isolates it was possible to amplify KSα and KSβ genes from environmental isolates from each crop site. This lack of correlation between direct and indirect isolation methods has been observed

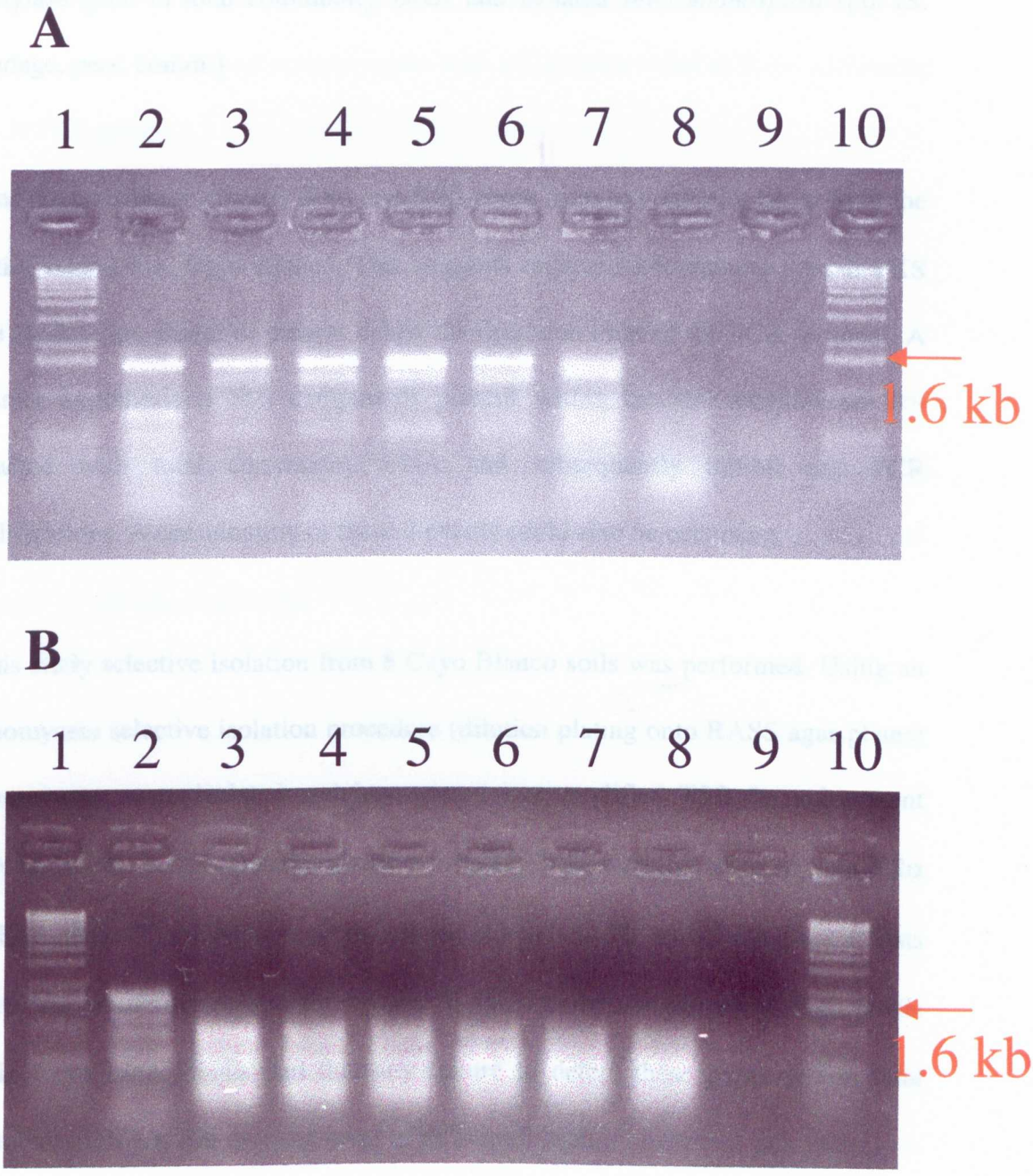


Figure3.5. (A) PCR amplification of α and β KS genes from a dilution series of *S. coelicolor* DNA. *S. coelicolor* DNA was diluted from 100 ng µl-1 to 1 pg µl-1 using ddH20 and 1 µl used per reaction. Lane 1. 1 kb marker, 2-8. DNA dilution series, 9. ddH2O, 10. 1 kb marker. **(B) PCR amplification of α and β KS genes from a dilution series of *S. coelicolor* DNA.** *S. coelicolor* DNA was diluted from 100 ng µl-1 to 1 pg µl-1 using total community DNA from a Cayo Blanco site and 1 µl used per reaction. Lane 1. 1 kb marker, 2-8. DNA dilution series, 9. ddH2O, 10. 1 kb marker.

in previous studies examining molecular detection of the gentamicin resistance methylase gene in total community DNA and isolated *Micromonospora* spp. (S. Standage, pers. comm.)

It was not possible to amplify KSα and KSβ genes from any samples taken from the pristine island site, Cayo Blanco. This suggests organisms containing type II PKS gene clusters are absent or present below the detection limit of the PCR protocol. A separate hypothesis is that compounds present within the soil samples are co-extracted with total community DNA and subsequently inhibit any PCR amplifications. A combination of these 2 events could also be occurring.

In this study selective isolation from 8 Cayo Blanco soils was performed. Using an actinomycete selective isolation procedure (dilution plating onto RASS agar plates) actinomycetes were isolated and innoculated into modified TSB for subsequent DNA isolation. Out of 50 actinomycetes isolated 90 % were found to contain KSα and KSβ genes. These bacteria were isolated at levels of 10^5 to 10^7 cfu g^{-1} soil. This shows that actinomyctes containing type II PKS genes are present at high levels within Cayo Blanco soils and that any failure to detect these genes within total community DNA is due to some level of PCR inhibition.

Evidence suggests that there is either a high population of actinomycetes in Cayo Blanco soils containing type II PKS gene clusters or a low diversity of actinomycetes with a high percentage of these bacteria containing the genes of interest. DGGE analysis of PCR products amplified using high G + C specific primers suggested a small population of actinomycetes in Cayo Blanco soils when

compared to high diversity soils (A.S. Anderson, pers. comm.). This data coupled with the high percentage of actinomycete isolates containing type II PKS genes suggests a low diversity of actinomcyetes with a high proportion of these containing type II PKS genes.

Total community DNA extracted from Cayo Blanco samples was purified by agarose gel electrophoresis and subsequent gel extraction. Once diluted to a working stock concentration (100 ng μl^{-1}) this DNA was subjected to PCR. It was not possible to amplify KSα and KSβ genes from this purified DNA. This suggests PCR inhibitory compounds may be closely linked to DNA molecules and are not removed by gel electrophoresis.

PCR was carried out on a dilution series of *S. coelicolor* A3(2) DNA starting with a concentration of 100 ng μl^{-1}. This DNA was diluted in a ten fold dilution series using ddH_20 from 100 ng μl^{-1} to 0.1 pg μl^{-1}. PCR product was detected in this experiment at a level down to 10 pg of DNA per reaction. This experiment was repeated using total community DNA extracted from a Cayo Blanco fir site to dilute *S. coelicolor* A3(2) DNA. This resulted in a PCR product only being visible in the undiluted *S. coelicolor* A3(2) DNA sample. This suggests that some compound co-extracted with total community DNA is capable of inhibiting polymerase reactions.

It was possible to amplify 16S rDNA genes using high G + C specific primers from all Cayo Blanco sites. This suggests that some PCR reactions are more robust than others although the concentration of target is higher for 16S rDNA genes. This is based on evidence that many bacteria contain multiple rRNA operons for example

Thermonospora chromogena has been found to contain six rRNA operons (Yap *et al.*, 1999). Actinomycete isolates containing type II PKS were found at between 10^5 to 10^7 cfu g^{-1} soil which suggests that the level of PKS genes were above the detection level of the PCR protocol and that the lack of PCR amplification was due to inhibition and not problems in the detection limit.

These results suggest that care should be taken when interpreting data from molecular screening experiments. In can be seen from the above data that results may indicate false negatives and that soils from which no amplification product can be observed by PCR may contain a large population of organisms containing the particular genes of interest. It would appear in the above Cayo Blanco example that co-extraction of PCR inhibitory compounds is the reason for the inability to amplify α and β KS from the total community DNA samples. In other samples though there may be other potential explanations for example incomplete cell lysis during DNA extraction techniques. It would be appear that perhaps the best approach would be to use a number of different cells lysis and DNA extraction techniques coupled with various DNA purification protocols to increase the chances of cell lysis and recovering DNA sufficiently pure for PCR amplification. This may prove impractical though for studies where high sample number is required.

3.8.1. Contribution to the field of study

- The first study to obtain homologous α and β KS gene pairs directly from total community DNA.

- The first specific study to use molecular detection techniques for the amplification of type II PKS genes as a method to investigate the distribution of these genes in terrestrial environments.
- One of the first studies to compare the differences between the direct molecular detection of antibiotic biosynthetic genes and the isolation of bacteria from the soil environment that contain these genes.

3.8.2. Future work

- Repeat the molecular detection of type II PKS genes using community DNA from different environments.
- To repeat the molecular detection in community DNA samples using other genes involved in polyketide biosynthesis.
- To develop RT-PCR methods with polyketide biosynthetic genes to investigate the activity of polyketide-producing actinomycetes in natural microbial communities.
- To develop molecular methods for the detection of specific type II polyketide genes.

Chapter 4

Molecular diversity of actinomycetes isolated from Cuban soil based on sequence analysis of 16S rDNA, *recA* and type II PKS genes

4.1. Introduction

The use of bacterial morphology and phenotypic traits for microbial identification often provides indistinct results. Although useful in determining subgeneric relationships they do not fully reflect the phylogeny of *Streptomyces* spp. Phylogenetic analysis of nucleic acid sequences can be used to determine these relationships (Olsen and Woese, 1993; Ludwig *et al.*, 1993). Phylogenetic analysis of nucleotide sequences is based on several assumptions, including that the evolution of the genetic sequence reflects strain evolution and that the rate of evolution is constant between groups. Studies have shown this not to be the case including analysis of 16S rDNA genes from *Vibrio cholera* indicating recombination between 16S rDNA genes leading to the loss of a marker restriction site (Lan and Reeves, 1992).

Small subunit ribosomal RNAs (SS-rRNA) have been extensively studied as phylogenetic marker molecules (Ludwig and Schleifer, 1994). The 5S and 16S rRNAs have been used for most rRNA based phylogenetic characterisations although the 16S rRNA is used more commonly since the development of more efficient DNA cloning and sequencing protocols have overcome sequencing constraints (Olsen *et al.*, 1986). SS-rRNAs exhibit a high degree of functional constancy and regions of sequence change at different rates allowing phylogenetic relationships to be measured (Woese, 1987). In the majority of bacterial species studied the genes encoding the SS-rRNAs are arranged in the order 16S-23S-5S. Analysis of the differences of the intergenic spacer regions between rRNA genes can also be used to distinguish bacterial relationships. Hain *et al.* (1997) used 16S-23S

rDNA intergenic spacer analysis to discriminate streptomycete strains using *S. albidoflavus* as an example. High resolution polyacrylamide gel electrophoresis of dye labelled PCR products was used as a rapid method to determine spacer size and number (Hain *et al.*, 1997).

It has become apparent though that reliance on one phylogenetic marker may lead to inaccurate results. The use of SS-rRNAs was based on the assumption that multiple SS-rRNA operons within an organism are identical or nearly identical and are not subjected to horizontal gene transfer. The actinomycete *Thermobispora bispora* was found to contain 2 distinct types of 16S rRNA genes by sequence analysis. The 2 types of 16S rRNA were found to differ at 98 nucleotide positions (6.4 % of total nucleotides) together with 6 regions of deletion-insertions. No difference was observed in evolutionary invariable or rarely variable nucleotides suggesting that both genes are transcriptionally active (Wang *et al.*, 1997). Sequence analysis of *Theromonospora chromogena* revealed six rRNA operons, 5 of which appear nearly identical. The sixth exhibits high levels of sequence variation in the coding regions of the 16S and 23S of approximately 6-10 %. This operon closely resembles an rRNA operon from another thermophilic actinomycete *Thermobispora bispora* suggesting possible acquisition via horizontal gene transfer (Yap *et al.*, 1999). Other candidates have been proposed to compare with SS-rRNA molecules. These include amino acid and DNA sequences for GroEL, ATP-ase-β-subunit, RNA polymerases and RecA (Eisen, 1995). Ludwig *et al.* (1993) analysed *tuf* (EF-tu protein chain) and *atpD* (ATP-ase-β-subunit) genes from different bacteria and showed similar phylogenetic relationships to those obtained with 16S rDNA phylogeny.

The RecA protein of *Escherichia coli* is a small (352 aa) protein with roles in homologous DNA recombination, SOS induction and DNA damage induced mutagenesis (Kowalczykowski *et al.*, 1994). The *recA* gene is one of the most highly conserved genetic elements to be found among the prokaryotes although it appears that the protein structure is more conserved than the DNA sequence which is altered to reflect the general G + C content of the species (Miller, 1992).

Lloyd and Sharp (1993) examined the DNA sequences of the *recA* gene from 25 bacterial strains. Phylogenetic analysis showed concordance between these data and other sequences based on largely 16S rRNA. Eisen (1995) analysed 65 complete RecA protein sequences using multiple parsimony and distance matrix methods. Phylogenetic trees based on these RecA sequences showed consistent subdivisions corresponding to a number of major bacterial groups and were found to be highly congruent to phylogenetic trees generated for SS-rRNA sequences. Sequences from the high GC species analysed in this study (3 *Streptomyces* spp., 2 *Mycobacterium* spp. and one *Corynebacterium* sp.) clustered together with the same branching pattern in both RecA and SS-rRNA sequence analysis. This suggests RecA sequence data may be useful in determining phylogenetic relationships within actinomycetes.

4.1.1. Aims

- To identify actinomycete isolates containing type II PKS genes.

- To establish the phylogenetic relationship of these isolates based on 16S rDNA and *recA* sequences.

- To compare these phylogenetic relationships with genetic diversity seen within antibiotic production genes.

4.2. Actinomycete isolation from Cuban soil

Selective isolation of actinomycetes was achieved by air drying 9 Cuban crop soils and subsequent dilution plating onto RASS agar containing cycloheximide (50 µl ml^{-1}) and nystatin (50 µl ml^{-1}) to inhibit fungal growth (2.6). Plates were incubated at 30°C for 14 d and colonies picked and enumerated. Actinomycetes were isolated from all soils at levels of 10^7 cfu g^{-1} of soil.

4.2.1. Total bacterial isolation from Cuban soil

Total bacterial isolation was achieved by plating a dilution series of Cuban crop soils resuspended in ¼ strength Ringers solution onto NA medium containing cycloheximide (50 µl ml^{-1}) and nystatin (50 µl ml^{-1}). Plates were incubated at 30°C for 7 d and colonies enumerated (2.6). Using this isolation procedure bacterial counts were found to be approximately 10^8 cfu g^{-1} of soil, approximately 1 log higher than total actinomycete counts (Figure 4.1).

4.3. Screening isolates for the presence of type II PKS genes

Actinomycete isolates were cultured in modified TSB and incubated at 30°C for 5 d on an orbital shaker. DNA was extracted from these strains using a small scale DNA extraction method (2.12.) and diluted to a working stock of 100 ng μl^{-1}. Polyketide (PKS 1 and PKS 2) specific primers were used to screen this isolate DNA for the presence of α and β KS genes. The presence of α and β KS genes was confirmed using semi-nested PCR (SN1 and SN2). Out of 100 actinomycete isolates screened 26 were found to contain α and β KS genes. It was possible to isolate actinomycetes containing α and β KS genes from all soils tested. Out of 9 soils tested it was possible to amplify type II PKS genes from total community DNA of only 4 soils (Table 3.5). This suggests that either actinomycetes containing type II PKS genes are present at levels below the detection limits of the PCR protocol when applied to total community DNA or that compounds co-extracted with total community DNA inhibit the PCR reaction as discussed in 3.8.

4.4. Amplification of 16S rDNA genes by PCR

16S rDNA genes were amplified from actinomycete isolates using universal bacterial PCR primers (pA and pH; Edwards *et al.*, 1989). These primers flank the extreme 5' and 3' part of the 16S rRNA gene allowing PCR amplification of the whole gene. This gave a PCR product of 1534 bp as shown in figure 4.2.

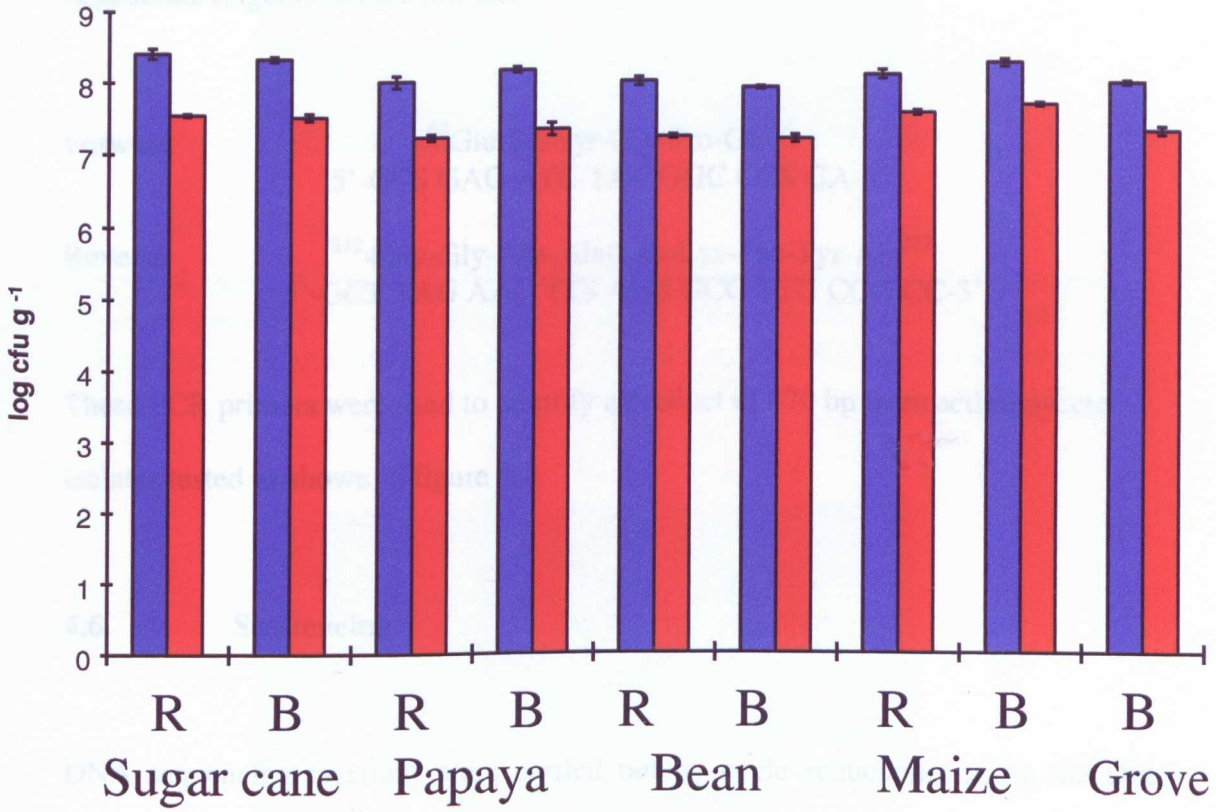


Figure 4.1. Total bacterial and actinomycete counts from 9 Cuban crop soils. Total bacterial counts were prepared on NA and actinomycete counts on RASS. Samples were taken from 5 Cuban crop sites and counts from rhizosphere (R) and bulk (B) soil samples were elucidated.

4.5. Amplification of *recA* genes by PCR

Nuβbaumer and Wohlleben (1994) described a method for amplifying the *recA* gene of *Streptomyces lividans* TK24. PCR primers were designed from an alignment of amino acid sequences of *recA* genes from 14 Gram negative bacteria and 2 Cyanobacteria. This showed 2 highly conserved regions giving the basis for 2 degenerate oligonucleotide primers.

Forward ^{64}Glu-Ile-Tyr-Gly-Pro-Gln^{70}
5'-GTS GAG ATC TAC GGC CCS GA-3'

Reverse ^{212}4Gly-Gly-Asn-Ala-Leu-Lys-Phe-Tyr-Ala^{222}
5'-GCT TAG AAC TTS AGS GCG TTG CCG CC-3'

These PCR primers were used to amplify a product of 470 bp from actinomycete isolates tested as shown in figure 4.3.

4.6. Sequencing

DNA sequencing reactions were carried out by cycle sequencing using the Dye Terminator Kit of PE Applied Biosystems (Warrington, Cheshire, UK). Sequencing reactions were carried out as per the manufacturers instructions (2.15). Duplicate sequencing reactions were carried out and analysed using Blast 2.0 (http://www.ncbi.nlm.nih.gov). Correct sequences were aligned in ClustalW (http://www2.ebi.ac.uk/clustalW) and this alignment used to construct consensus sequences for each sample.

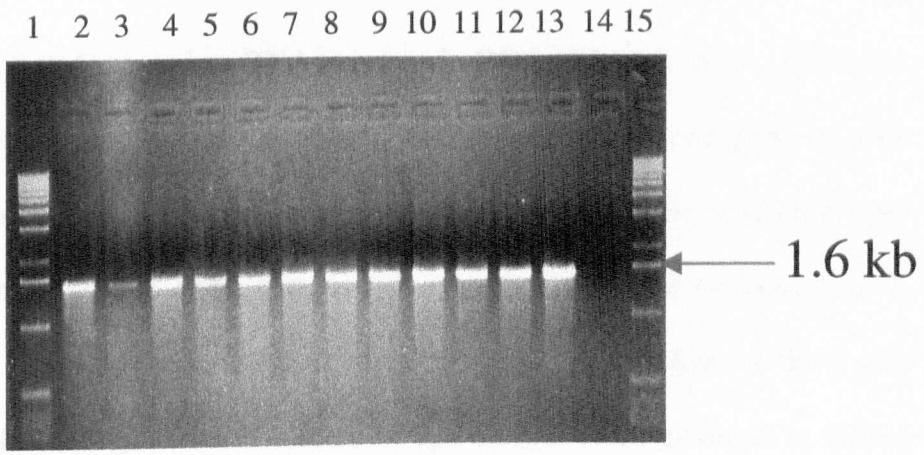


Figure 4.2. Amplification of 16S rDNA genes from actinomycete isolates using universal bacterial PCR primers (pA and pH, Edwards *et al.*, 1989). Lane 1. 1 kb marker, 2-12. Cuban isolates, 13. *S. coelicolor* DNA, 14. ddH_2O. 15. 1 kb marker.

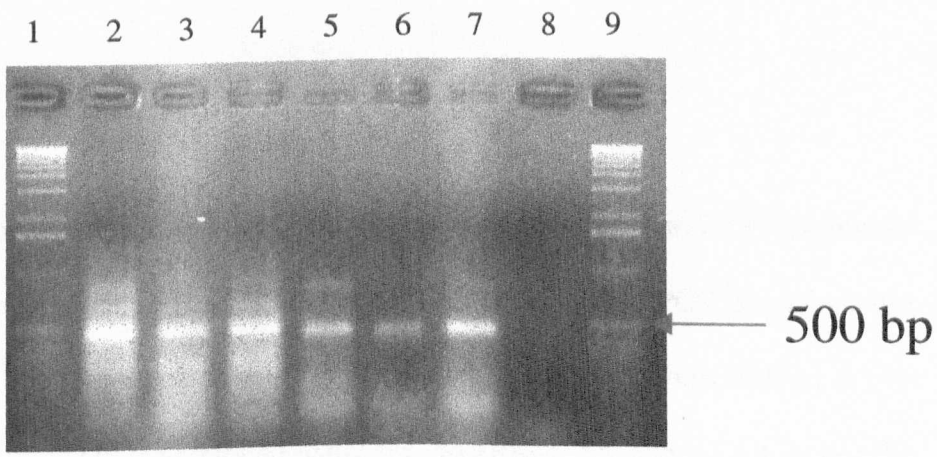


Figure 4.3. Amplification of *recA* genes from actinomycete isolates using degenerate oligonucleotide primers (Nuβbaumer and Wohlleben, 1994). Lane 1. 1 kb marker, 2-6. Cuban isolates, 7. *S. coelicolor* DNA, 8. ddH_2O. 9. 1 kb ladder.

4.6.1. 16S rDNA and *rec*A sequencing

Products amplified from isolate DNA using 16S rDNA (Edwards *et al.*, 1989) and *rec*A (Nuβbaumer and Wohlleben, 1984) specific oligonucleotide primers were identified using agarose gel electrophoresis. These products were extracted from agarose gels and purified using Qiagen gel extraction (Qiagen) or Geneclean II (Bio101) kits as per the manufacturers instructions (2.8.2). Samples of these gel-extracted products (5 µl) were subjected to agarose gel electrophoresis to ensure sufficient concentration of PCR amplified DNA. Aliquots of these purified products (2.5 µl) were used in subsequent sequencing reactions (2.15). Samples were sequenced using the forward primers pA and RecA1.

4.6.2. Type II PKS gene sequencing

Type II PKS genes were amplified from Cuban isolates using α and β ketosynthase specific oligonucleotide primers (PKS1 and PKS2, This study). PCR products were gel extracted (2.8.2) and quantified using spectrophotometer readings (OD_{260}, 2.11). Purified PCR products were cloned using the Invitrogen TA cloning kit as per the manufacturers instructions (Invitrogen, 2.14). Transformants were picked and grown up in LB broth containing ampicillin (50 µg ml^{-1}) and plasmid DNA prepared using a QIAprep Spin Plasmid Kit (QIAGEN). Plasmid DNA was quantified using spectrophotometer readings (OD_{260}, 2.11) and approximately 800 ng used for subsequent sequencing reactions. Plasmids were sequenced using T7 and M13 forward and reverse primers (Invitrogen). Type II PKS genes were cloned

and sequenced due to the consensus nature of the oligonucleotide primers used for the PCR amplification. Direct sequencing of PCR amplified products proved unsuccessful. This was believed to be due to unsuccessful binding of consensus oligonucleotide primers to the plasmid DNA during the sequencing reaction. Sequencing of cloned PCR products using internal primer binding sites on the pCR2.1 vector backbone proved successful.

4.7. Phylogenetic analysis of sequence data

DNA sequences were confirmed by comparison with published sequences using the Blast sequence database (http://www.ncbi.nlm.nih.gov) and consensus sequences produced from duplicate samples using the ClustalW alignment package (http://www2.ebi.ac.uk/clustalW). Dendrograms were constructed using the PHYLIP software package (version 3.5) and trees drawn using Treeview. Bootstrap values from 100 replicates were calculated using PHYLIP as indicated by Felsenstein (1985, 2.16).

4.7.1. Analysis of 16S rDNA gene sequences

Universal 16S rDNA oligonucleotide primers were used to amplify 16S rDNA genes from 23 actinomycete isolates and these were sequenced, in duplicate, using the forward primer (pA). The first 350 bp of these sequences taken from the start of

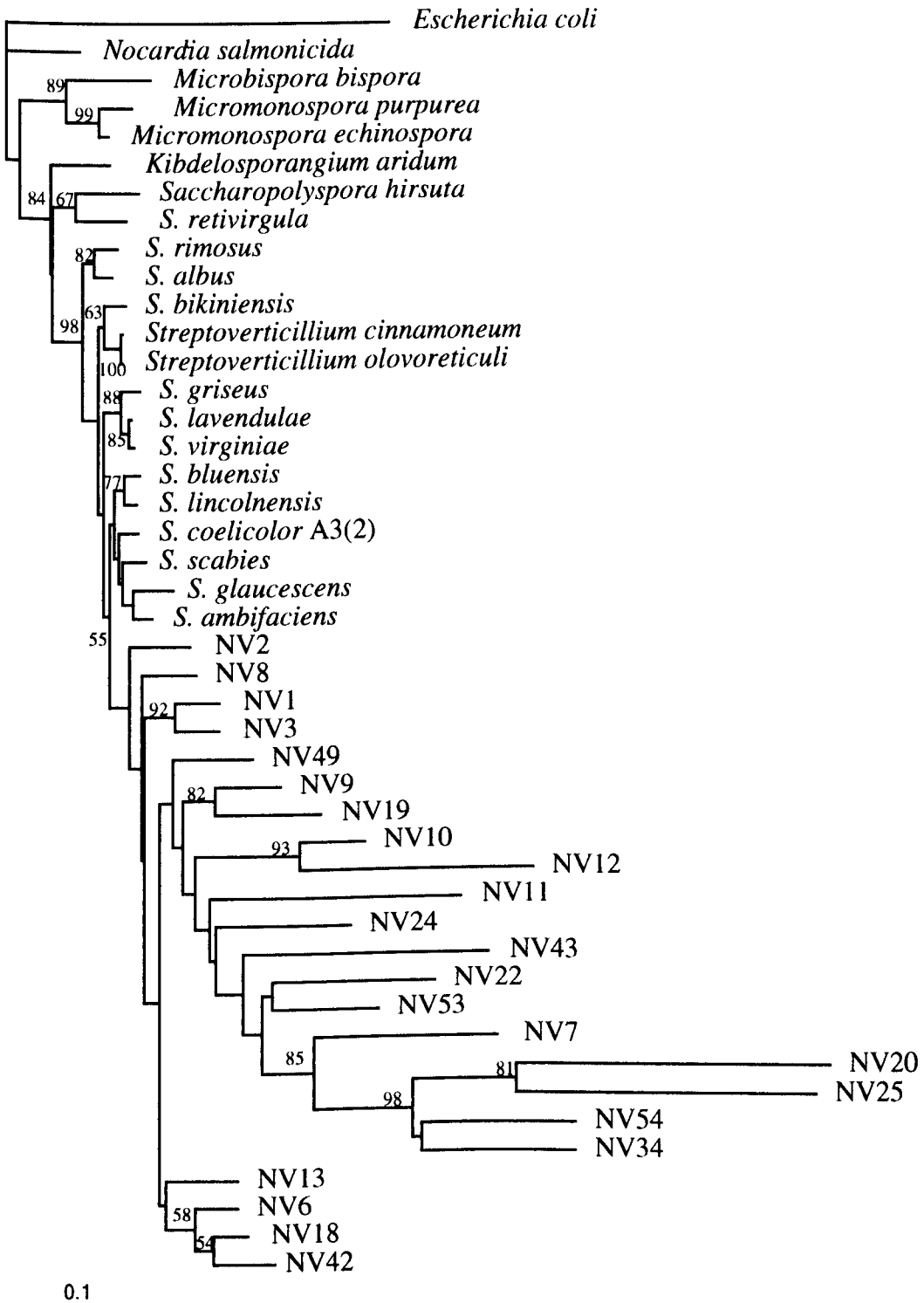


Figure 4.4. Phylogenetic tree of 16S rDNA sequences including actinomycete isolates and actinomcycete type strains based on partial sequence (350 bp). Dendrogram deriverd using a Neighbor-Joining method. Dendrogram bootstrapped 100 times, with values over 50 % shown. *E. coli* used as an outgroup.

the priming site were used for further phylogenetic analysis. These 23 sequences along with 21 actinomycete type strains were compared as indicated in 4.7. using the *E. coli* 16S rRNA gene as an outgroup (Figure 4.4). The resulting dendrogram showed the Cuban isolates clustering within the streptomycete group. These isolates formed a separate group when compared to streptomycete type strains. Branch length analysis revealed that although clustering together these isolates appeared diverse when compared to each other.

4.7.2. Analysis of *recA* gene sequences

Primers designed to amplify *recA* genes were used to amplify *recA* genes from 12 actinomycete isolates. Sequencing of these products was carried out in duplicate using the forward primer with 400 bp from the start of the priming site used for phylogenetic analysis (4.7). Phylogenetic analysis of these sequences in comparison with available type strain sequences showed these isolates to form a cluster within the streptomycete group (Figure 4.5). Branch length analysis of these isolates suggested them to be less diverse than those observed using 16S rDNA genes as a phylogenetic marker.

4.7.3. Analysis of type II PKS gene sequences

PCR products from 4 actinomycete isolates were cloned and duplicate clones sequenced using the T7 sequencing primer. Sequence analysis was carried out on 400 bp from the start of the PKS primer site. All 4 sequences were found to show a

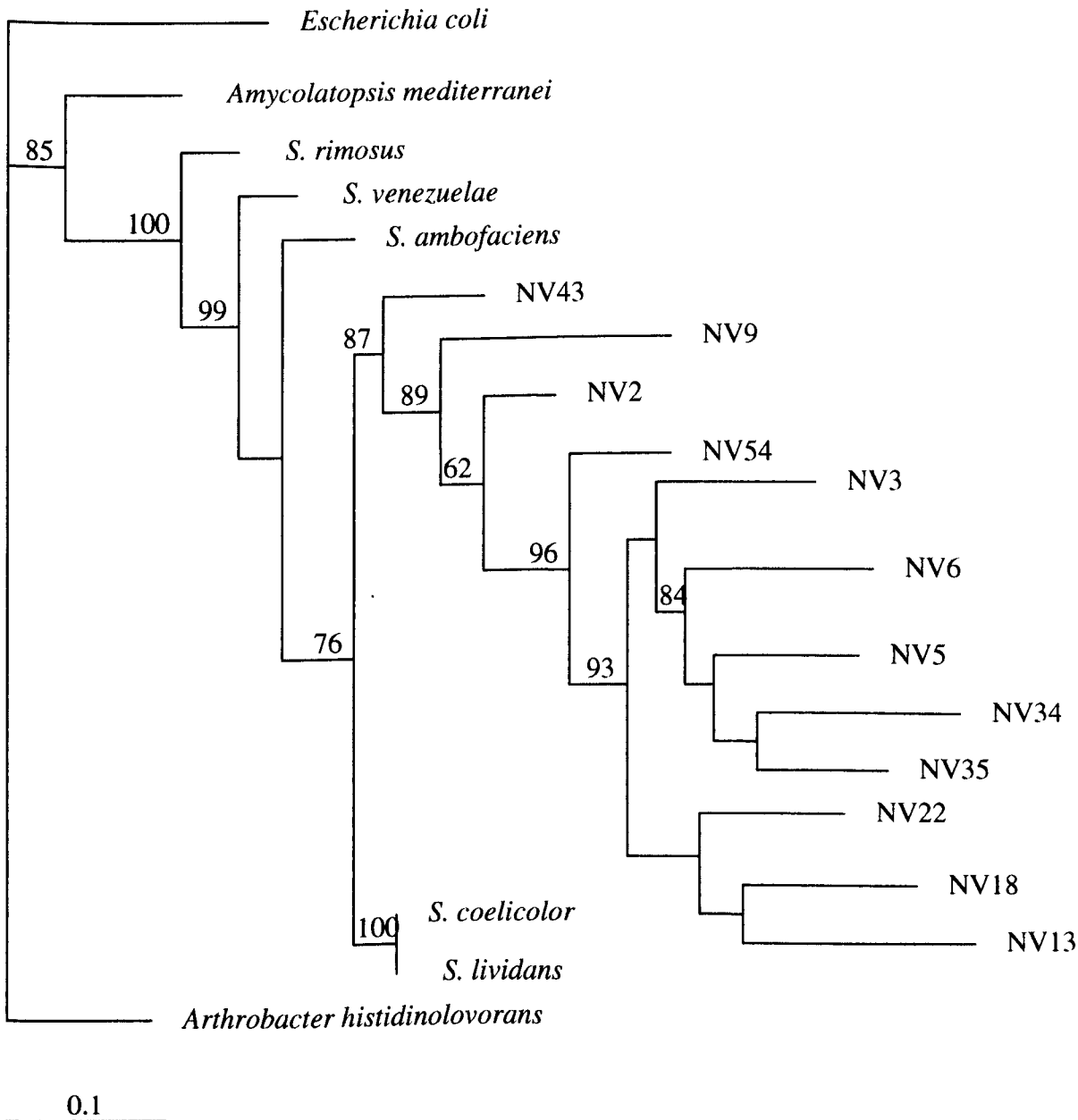


Figure 4.5. Phylogenetic tree of *recA* sequences including actinomycete isolates and actinomycete type strains based on partial sequence (400 bp). Dendrogram deriverd using a Neighbor-joining method. Dendrogram bootstrapped 100 times, with values over 50 % shown. *E. coli* used as an outgroup.

high similarity (84-90 %) at a nucleotide level to KSα genes from actinomycete type strain sequences. The amino acid identity was found to be between 51-71 % and the similarity between 66-85 %. At the nucleotide level three sequences resembled type II PKS genes responsible for the production of polyketide antibiotics and 1 resembled a spore pigment gene. This trend was also observed at the amino acid sequence level. One isolate NV19 showed similarity at the nucleotide level to a KSα gene isolated from *S. albus*. This gene is from an unidentified PKS cluster and homology to known KSα genes from other type strains is the only evidence for a polyketide product. Other nucleotide matches to this gene suggest it may be responsible for the production of a polyketide antibiotic as opposed to a spore pigment.

Cuban Isolate	% amino acid similarity	Polyketide	% nucleotide similarity	Polyketide
NV9	*S. halstedii* I : 51 %, S : 69 %	spore pigment	*S. halstedii*, 88 %	spore pigment
NV19	*S. venezuelae* I : 69 %, S : 85 %	jadomycin	*S. albus*, 84 %	KSα
NV22	*S. venezuelae* I : 71 %, S : 84 %	jadomycin	*S. venezuelae*, 86 %	jadomycin
NV49	*S. venezuelae* I : 55 %, S : 66 %	jadomycin	*S. argillaceus*, 90 %	*mithramycin*

Table 4.1. Percentage similarity between Cuban isolates containing a type II PKS KSα compared to actinomycete type strains at both the amino acid and nucleotide level. Isolate sequences (400 bp) were compared to nucleotide and amino acid sequences using the BLAST database at the NCBI (blastn and blastx). For amino acid sequences the identity (I) and similarity (S) are quoted.

4.8. Discussion

Actinomycetes were isolated from 9 Cuban crop soils using a selective isolation method. This was repeated using NA media for total bacterial counts. These were found to be approximately 10^8 cfu g^{-1} soil compared with actinomycete counts of

approximately 10^7 cfu g^{-1} of soil. The same counts were observed across all crop soils tested with no differences seen between bulk and rhizosphere samples. This suggests that actinomycete communities are at similar levels in both rhizosphere and bulk soil although the diversity present within those communities may be different.

100 actinomycetes were picked and tested for the presence of type II PKS genes using the PCR protocol developed in Chapter 3. Out of these isolates 26 were found to contain α and β KS genes. It was possible to amplify 16S rDNA genes from all 100 strains using both universal and high G + C specific primers. This suggests that the isolate DNA was of sufficient purity in all samples for PCR amplification. All PCR reactions were carried out in triplicate to avoid false negatives. The results suggest that compared with other soils for example, Cayo Blanco, the numbers of actinomycetes containing type II PKS genes are quite low.

PCR products amplified from Cuban isolates using 16S rRNA specific primers were sequenced using the forward primer (pA), and confirmed using the BLAST database at the NCBI. Partial sequences (350 bp) were used for phylogenetic analysis and compared to published actinomycete 16S rDNA sequences. All Cuban isolates were found to group with *Streptomyces* spp. although they grouped together to form a separate cluster. This may suggest the presence of a novel group of streptomycetes within the Cuban soils studied. Despite clustering together, a number of these isolates showed long branch lengths suggesting a high diversity within this novel group.

All 16S rRNAs have a common three-dimensional structure with primary structures and variable regions identified (Woese, 1987). Some regions are unique at the genus level (Salama *et al.*, 1991) and more variable regions can be used for assigning groups to lower taxanomic levels (Klijn *et al.*, 1991). Stackebrandt *et al.* (1992) identified 3 regions that vary within the 16S rRNA genes of streptomycetes; α (nt 982-998, *S. ambofaciens* nomenclature) and β (nt 1102-1122) which are suitable for resolution at the genus level. The third region γ (nt 150-200) is the most variable and can be used for species specific phylogeny. The region of the 16S rDNA genes sequenced in this study included the γ hypervariable region of the 16S rDNA gene. Stackerbrandt (1995) suggested this region may be unsuitable for phylogenetic analysis although sequence data from these Cuban isolates suggests that this region might be useful in discriminating intragenic relationships within streptomycetes.

PCR products amplified using *recA* specific primers were used to amplify *recA* genes from a number of Cuban isolates. These products were sequenced using the forward primer and partial sequence (400 bp) from the priming site used for phylogenetic analysis. Sequences were found to group within streptomycete type strain sequences, and in agreement with 16S rDNA sequences the isolates appeared to form a separate cluster group. This data is in agreement with previous studies on the study of RecA sequence data (Eison, 1995). Branch lengths were found to be shorter than those seen in previous phylogenetic analysis of 16S rDNA genes. This may be due to the fact that the *recA* sequences include the nucleotide region coding for the active site and the 16S rDNA sequence includes the hypervariable region. Due to the lack of available *recA* gene sequences it is unknown whether these

isolate sequences form their own cluster or whether further type strain sequencing would reveal other previously described *Streptomyces* spp. within this group.

Type II PKS genes amplified from Cuban isolates were found to be similar to previously published sequences. Similarity was seen to genes coding for both antibiotic and spore pigment polyketides. The 2 isolates studied by Seow *et al.* (1997) cultured from a Canadian soil both contained KSβ genes clustering within published antibiotic production genes. Metsä-Ketalä *et al.* (1999) screened a number of soil isolates using PCR primers specific to KSα genes. Isolates were found to cluster with published type II PKS sequences and to be representative of both antibiotic and spore pigment production genes. These isolates were found to show nucleotide sequence similarity between 84-94 % with amino acid sequence similarity between 60-99 % compared to published aromatic polyketide sequences. These data are in agreement with observations made in this study. Low sequence similarities suggest these sequences may belong to novel type II biosynthetic clusters. Amino acid sequence identity was found to be between 51-71 %. This result is expected given the structural variation of different polyketide compounds (Metsä-Ketalä *et al.*, 1999).

Selective isolation of soil bacteria can be a useful source of genes for novel bioactive products. One problem though is that isolation methods tend to enrich isolates that produce compounds that are already known and abundant in nature. Phylogenetic analysis of isolates can be used to isolate particular phylogenetic groups for further study. This can be coupled with the use of rapid PCR screening and gene sequencing to identify novel genes within large isolate collections. This

may provide a more rapid technique for the identification of novel biosynthetic gene clusters.

4.8.1. Contribution to the field of study

- Analysis of actinomycete isolates from Cuban crop soils containing type II PKS genes as determined by PCR using PKS specific oligonucleotide primers.
- Determined the ability to use *recA* sequence data as a method to ascertain the phylogenetic relationship between Cuban isolates and actinomycete type strains.
- Evaluated and compared the phylogenetic relationship between Cuban isolates using both 16S rDNA and *recA* sequence data.
- Provided sequence data from isolate KSα genes indicating distinct genes when compared to previously published sequences.

4.8.2. Future work

- Further characterisation of type II PKS genes to determine if the diversity seen in KSα genes is representative of the culturable population.
- Characterisation of KSβ genes from clusters containing sequenced KSα to determine the similarity between homologous KSαβ gene pairs.
- To sequence *recA* genes for described actinomycete type strains to help resolve inter and intrageneric relationships between genera.

Chapter 5

Diversity of type II polyketide synthase genes amplified directly from total community DNA

5.1. Introduction

It is estimated that >99 % of microorganisms observable in nature typically are not cultivated by using standard techniques (Amann *et al.*, 1995). It has been shown that one gram of soil may contain more than 10^{10} bacteria as determined by fluorescent microscopy (Torsvik *et al.*, 1990). The relative proportion of bacteria growing on agar plates varies from 0.1-1.0 % in pristine forest soils to 10 % in arable soils. This suggests that investigations based on bacterial isolates include only a small part of the total diversity present in the environment (Torsvik *et al.*, 1998). Recombinant DNA techniques and molecular phylogenetic methods have allowed the detection of organisms appearing in microbial communities without the need for cultivation (Hugenholtz and Pace, 1996). Application of these techniques to a number of diverse environments has shown the limited amount of known diversity and allowed access to novel and environmentally important diversity yet to be studied (Pace, 1997). Analysis of environmental DNA has shown substantial phylogenetic diversity with little or no representation when compared to previous culture-dependant studies. The wide distribution and abundance of these sequences suggests these organisms may contribute significantly to global chemical cycles and processes (Hugenholtz and Pace, 1986).

The study of SS-rRNAs and their respective genes has been used extensively to study the microbial diversity present in environmental samples. A number of methods are available for studying the diversity of sequences in these genes. Resolving PCR products amplified from 16SrRNA genes can be achieved using denaturing gradient gel electrophoresis (DGGE) to give information about changes

in community structure (Muyzer *et al.*, 1993). This can be combined with hybridisation using phylogenetic probes to give an understanding of the prominence of certain community members (Øvreås *et al.*, 1997). Cloning PCR products from total community DNA can elucidate the highest resolution. This allows comparison of the cultivated microbial community and the total microbial community. Sequencing of bacterial isolates and clones followed by subsequent phylogenetic analysis allows comparison of total and culturable microbial diversity.

Once cloned, gene libraries can be screened in a variety of ways. These include colony hybridisation using gene specific oligonucleotides or restriction digestion of plasmids. Probe specificity is important to avoid false positives during colony hybridisation (Head *et al.*, 1998). Automated sequencing allows rapid screening of large clone libraries and sequence data can be compared to a database. Phylogenetic analysis can then be carried out to allow the diversity of the microbial community to be determined with reference to previously published sequences (Maidak *et al.*, (1994).

Seow *et al.* (1997) used PCR primers to amplify KSβ genes from total community DNA. These products were cloned and 2 clones sequenced. Both sequences were found to cluster within known type II PKS gene clusters. Hopwood (1997) examined the phylogenetic relationship of amino acid sequences from type II PKS genes (Figure 5.1). KSα and KSβ genes were found to cluster separately suggesting

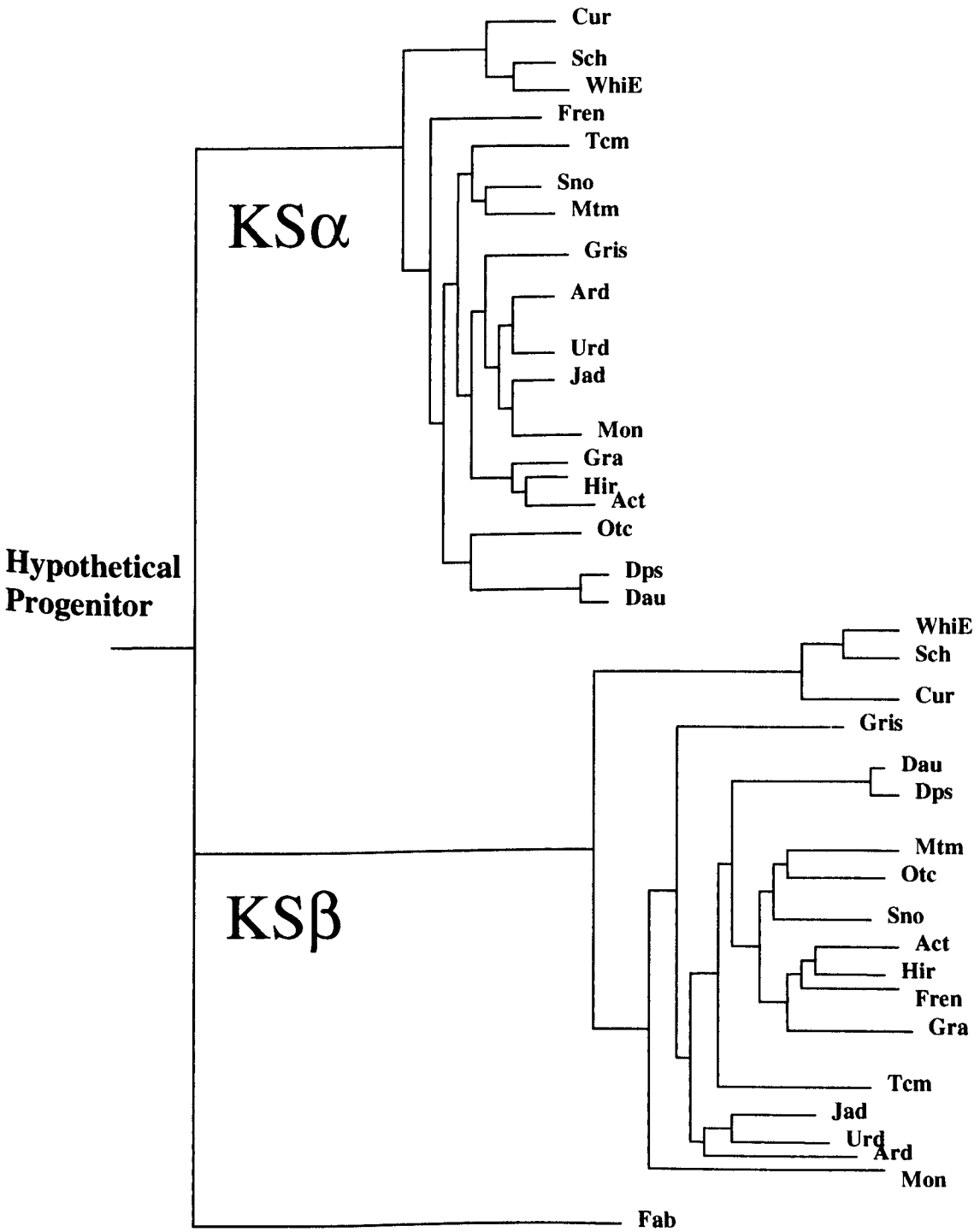


Figure 5.1 Phylogenetic tree of amino acid sequences of KSα and KSβ subunits of actinomycete type II PKSs. For key see Chapter 1, Table 1.4. The putative *S. glaucescens* FAS condensing enzyme (FAB) was used as an outgroup. Figure reproduced from Hopwood (1997).

different functions of these genes. Within these individual clusters the genes encoding polyketide spore pigments were found to be separate from those encoding the antibiotic (and other bioactive) polyketides. This suggests a separate line of descent for these 2 groups.

5.1.1. Aims

- To produce a representative clone library of type II PKS α and β KSs from total community DNA samples.

- To establish the phylogenetic relationship of these type II PKS sequences when compared to known type strain sequences.

- To analyse the diversity observed when comparing rhizosphere and bulk samples from a defined site.

5.2. Production of a representative clone library containing α and β KS genes.

PKS genes were amplified from total community DNA extracted from a Cuban crop site seeded with bean plants. This site had previously been shown to contain α and β KS genes as determined by PCR using polyketide specific primers (PKS 1 and PKS 2, Chapter 3). Both rhizosphere and bulk soil samples were available for this site (soil 205, 206). PCR products were identified by agarose gel electrophoresis and gel extracted using a Qiagen gel extraction kit (QIAGEN, 2.8.2). PCR products were

cloned into the pCR2.1 vector using an Invitrogen TA cloning kit and used to transform *E. coli* TOP 10F' competent cells. Positive transformants were picked, inoculated into LB (10 ml) containing ampicillin (50 μg ml^{-1}) and incubated overnight at 37°C. Plasmid preparation was carried out using a QIAprep Spin Plasmid Kit (QIAGEN). Transformants were stored by adding an equal volume of 50 % glycerol (1 ml) to an aliquot of overnight culture (1 ml) and stored at -20°C.

5.3. Sequencing

Clones were sequenced in duplicate using T7 (forward) and M13 (reverse) primers, which bind to internal regions within the vector backbone (2.15). Sequences were aligned against the PKS primers (PKS 1 and PKS 2) to determine the orientation of the PCR amplified insert within the vector using the sequence alignment programme ClustalW (http://www2ebi.ac.uk./clustalW). Partial nucleotide sequences (400 bp) starting at the PKS priming sites were then analysed and compared to known α and β KS genes using the BLAST database at the NCBI (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST).

5.4. Phylogenetic analysis of PKS sequences

5.4.1. PKS type strain sequences

Sequences of type II PKS α and β KS genes from actinomycete type stains were obtained using the Entrez nucleotide retrieval database at the NCBI (http://www.ncbi.nlm.nih.gov/Entrez/). Type strain sequences were aligned against the polyketide specific primers (PKS 1 and PKS 2, 5.3) and partial sequences (400 bp) of both α and β KS genes were analysed using the PHYLIP software package (2.16, Figure 5.2). Groupings were found to be consistent with phylogenetic relationships published relating to amino acid sequences by Hopwood (1997). The genes responsible for production of spore pigments (*cur*, *sch* and *whiE*) were found to group separately from those responsible for antibiotic production. This trend is seen in both the α and β KS genes although the KSβ genes responsible for spore pigment production appear more closely related to antibiotic biosynthetic genes than their KSα relative. High bootstrap values appear to give confidence to these observations. This suggests that these 400 bp from the PKS priming sites can be used to determine relationships between type II PKS genes.

5.4.2. Soil clone sequences

A number of soil clones were identified (50 from each soil type) and a random sample of these sequenced and compared to the type strain sequences analysed in 5.4.1. These sequences included α and β KS genes from both soil samples

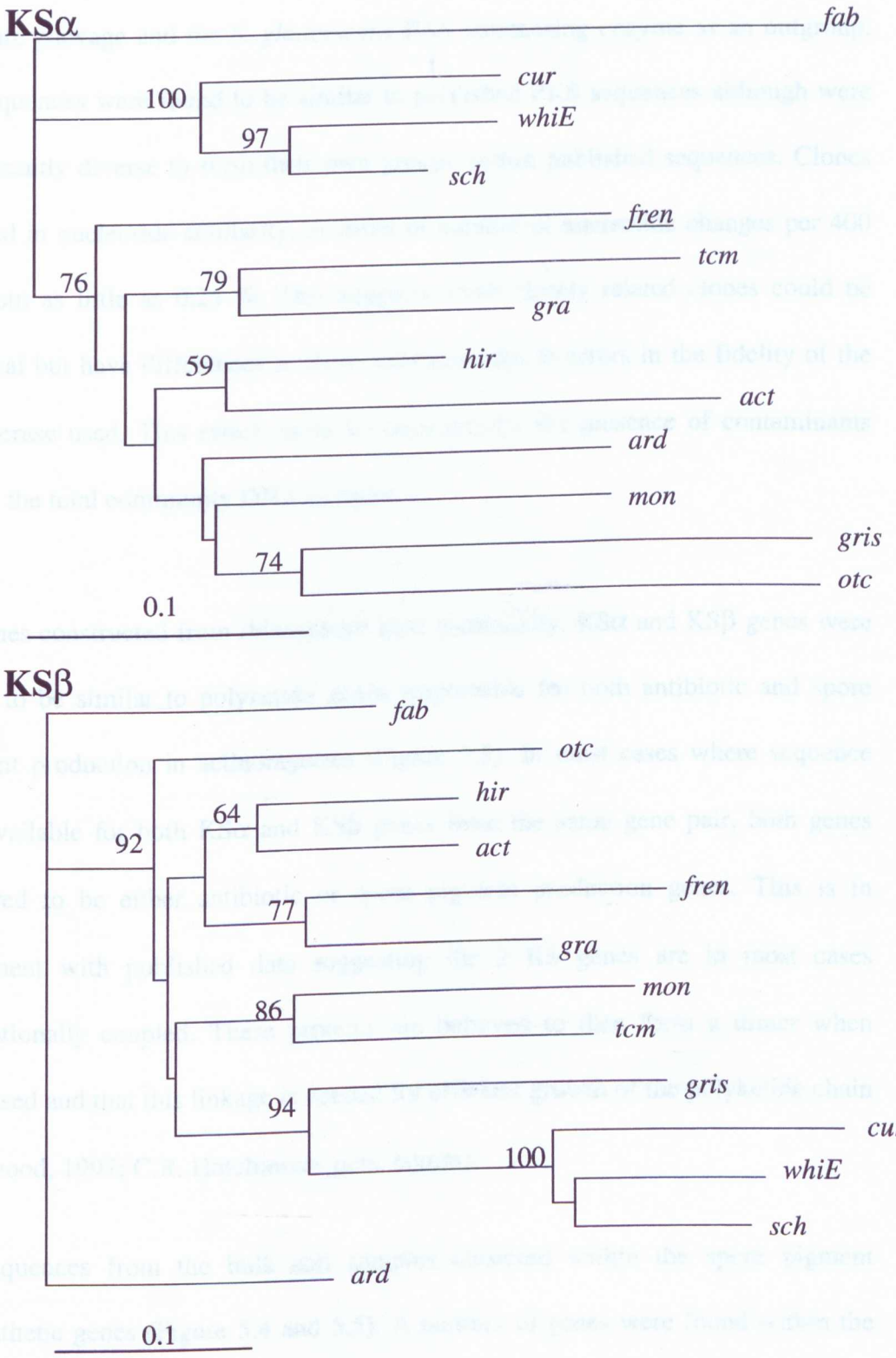


Figure 5.2. Phylogenetic trees of partial nucleotide sequences of KSα (A) and KSβ (B) subunits of actinomycete type II PKSs. Neighbor- joining trees were constructed using PHYLIP (version 3.5) and using the *S. glaucescens* FAS condensing enzyme as an outgroup. For key see Table 1.4.

(rhizosphere and bulk). Neighbour-joining trees were constructed using the PHYLIP software package and the *S. glaucescens* FAS condensing enzyme as an outgroup. All sequences were found to be similar to published PKS sequences although were significantly diverse to form their own groups within published sequences. Clones differed in nucleotide similarity, in terms of number of nucleotide changes per 400 bp, from as little as 0.25 %. This suggests some closely related clones could be identical but have differences in their sequences due to errors in the fidelity of the polymerase used. This effect could be increased by the presence of contaminants within the total community DNA samples.

In clones constructed from rhizosphere total community, KSα and KSβ genes were found to be similar to polyketide genes responsible for both antibiotic and spore pigment production in actinomycetes (Figure 5.3). In most cases where sequence was available for both KSα and KSβ genes from the same gene pair, both genes appeared to be either antibiotic or spore pigment production genes. This is in agreement with published data suggesting the 2 KS genes are in most cases translationally coupled. These proteins are believed to then form a dimer when expressed and that this linkage is needed for efficient growth of the polyketide chain (Hopwood, 1997; C.R. Hutchinson, pers. comm).

No sequences from the bulk soil samples clustered within the spore pigment biosynthetic genes (Figure 5.4 and 5.5). A number of genes were found within the rhizosphere clones that clustered within the spore pigment production genes. This suggests possibly that type II PKS genes responsible for the production of spore

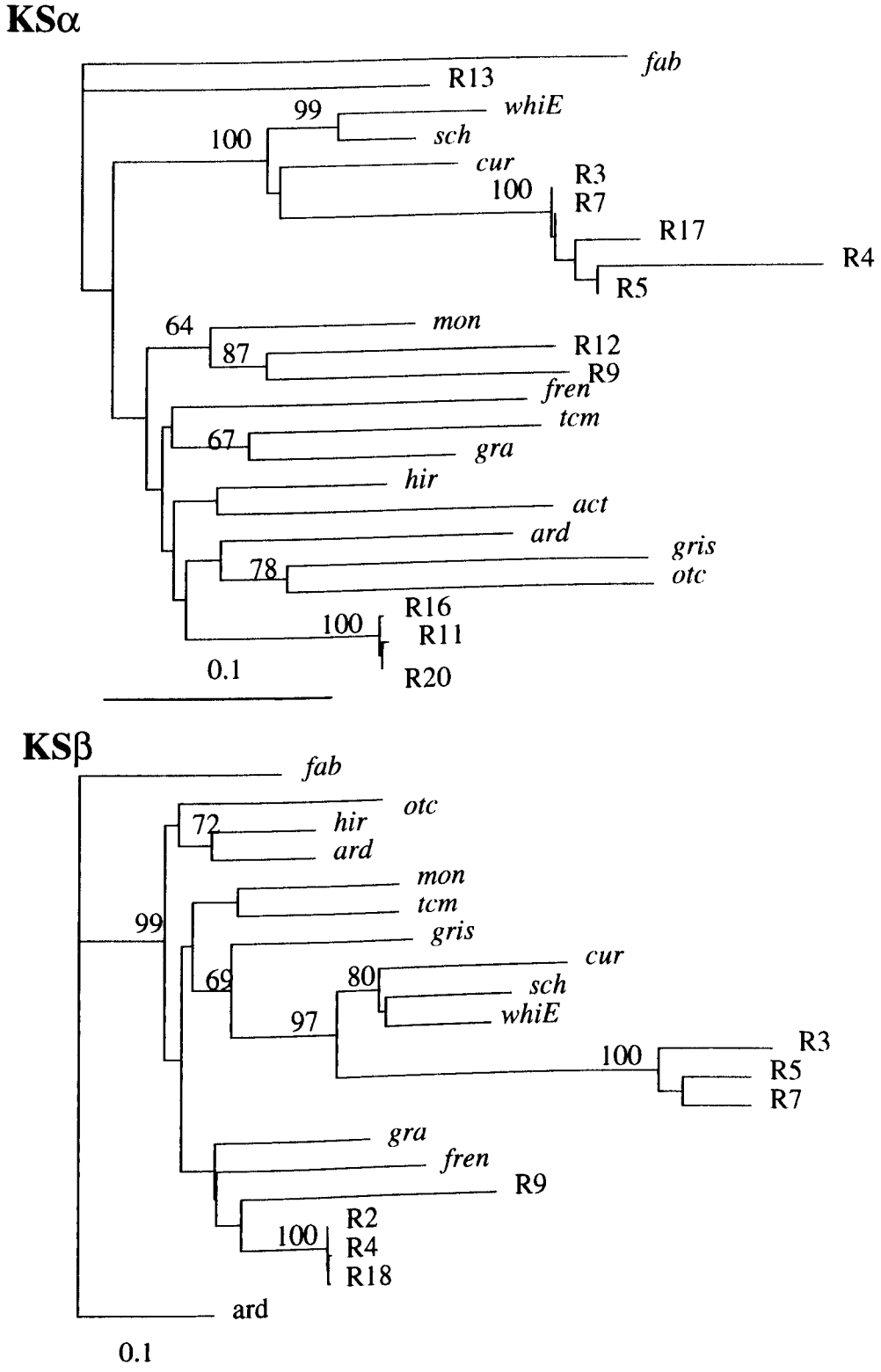


Figure 5.3. Neighbour-joining trees of partial nucleotide sequences (400 bp) of KSα and KSβ genes. Clone library sequences were obtained from rhizosphere (R) soil samples of bean crops. Trees were constructed using PHYLIP and the *S. glaucescens* FAS condensing enzyme as an outgroup. For key see table 1.4.

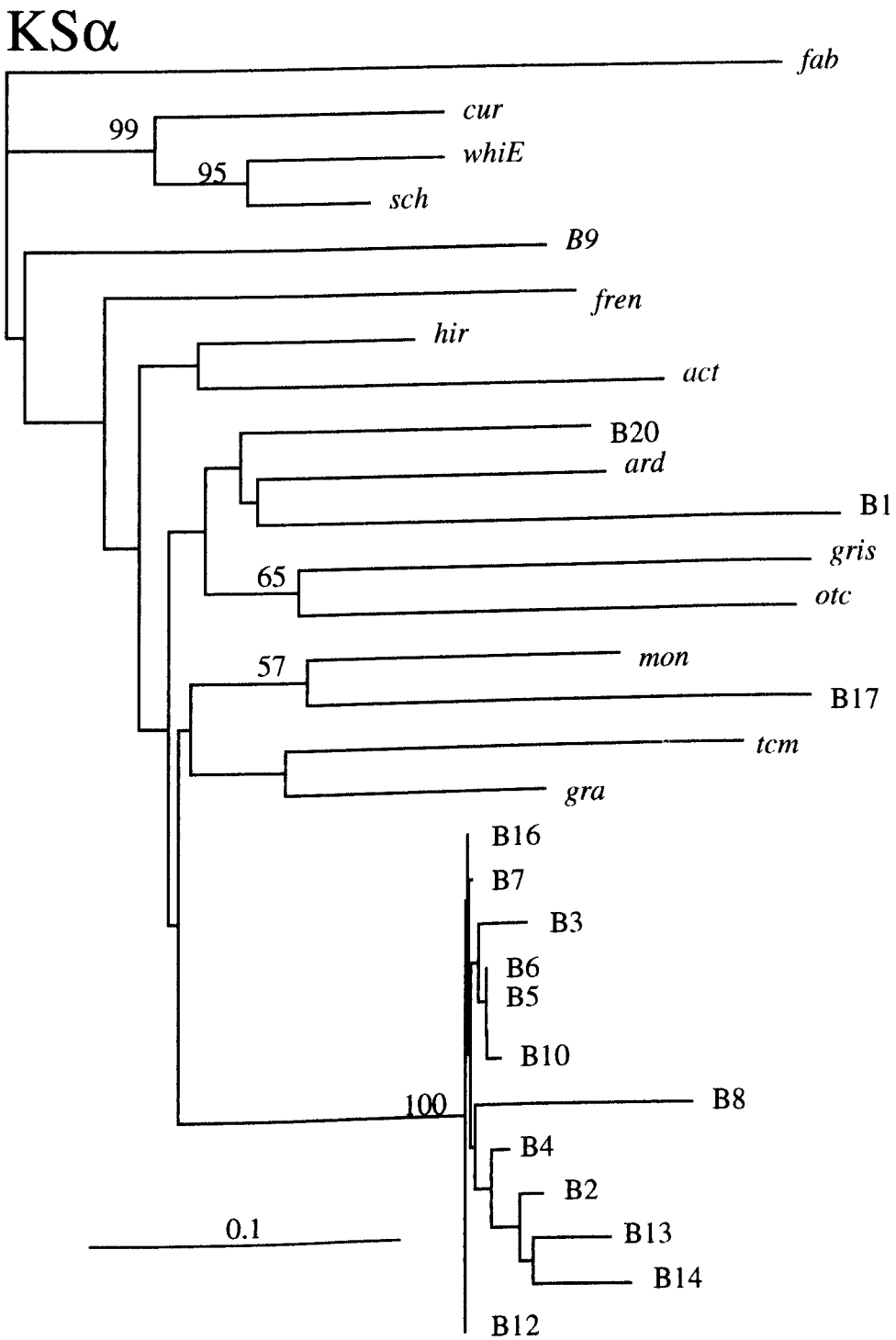


Figure 5.4. Neighbour-joining tree of partial nucleotide sequences (400 bp) of KSα genes obtained from bulk soil (B). Sequences were compared to actinomycete type strains using the PHYLIP software package and the *S. glaucescens* FAS condensing enzyme as an outgroup. For key see table 1.4.

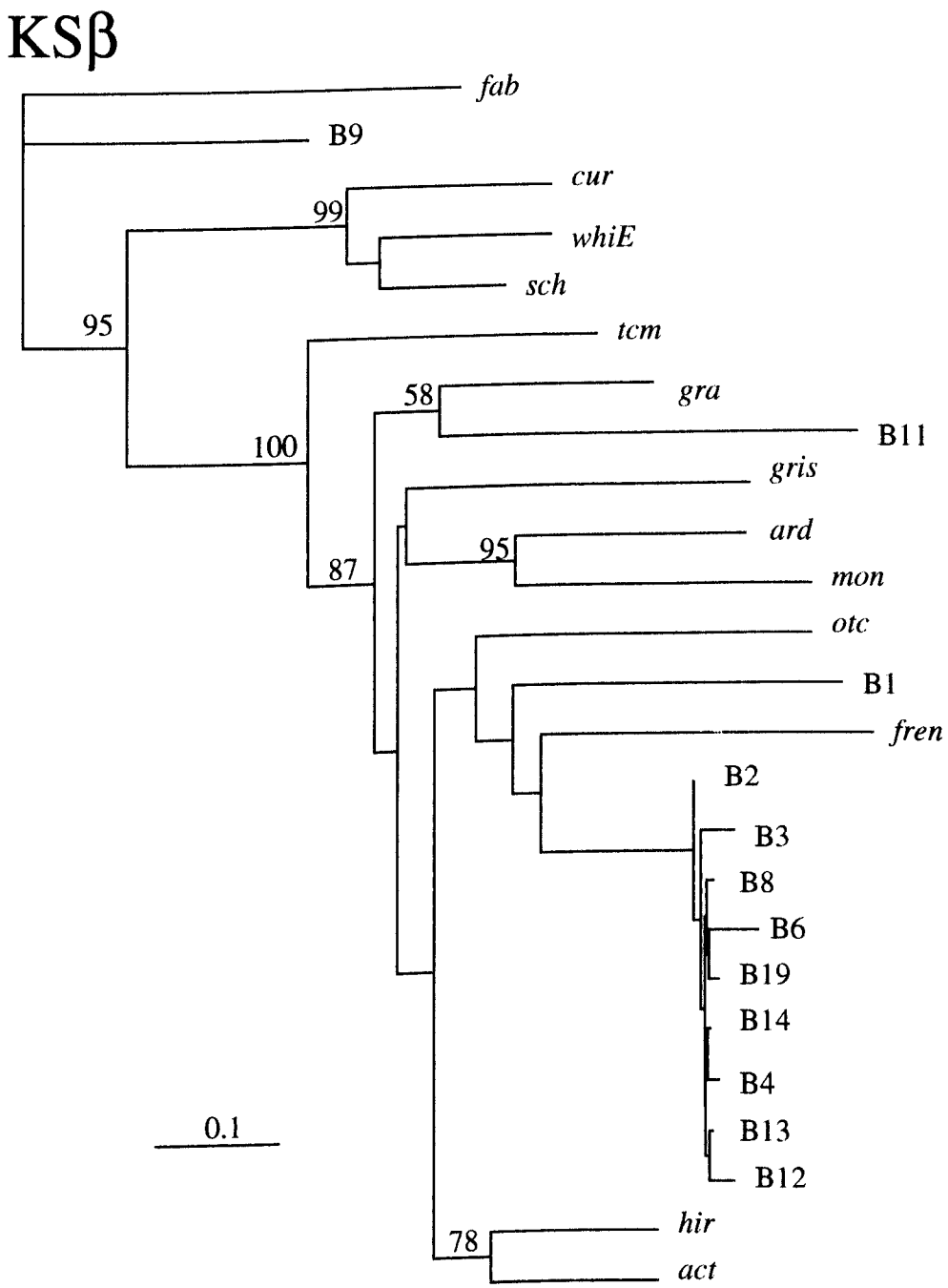


Figure 5.5. Neighbour-joining tree of partial nucleotide sequences (400 bp) of KSβ genes obtained from bulk soil (B). Sequences were compared to actinomycete type strains using the PHYLIP software package and the *S. glaucescens* FAS condensing enzyme as an outgroup. For key see table 1.4.

pigments maybe of greater significance to actinomycetes colonising the rhizosphere than those found in bulk soil.

5.5. Isolate screening

Isolates (200) were picked from actinomycete selective plates and inoculated into modified TSB. After subsequent DNA extraction and PCR, 36 % of isolates were found to contain type II PKS genes (Table 5.1).

Isolate site	+ve %
Total	36
Bean, Bulk	32
Bean, Rhizosphere	40

Table 5.1. Screening of 200 actinomycetes for the presence of type II PKS genes. Out of 200 actinomycetes isolated 36 % were found to contain type II PKS α and β KS genes as determined by PCR. A small difference was observed in the numbers of positive isolates when comparing the bulk and rhizosphere samples.

An equal number of isolates were picked from the bulk and rhizosphere samples. It was possible to amplify type II PKS genes from 40 % of rhizosphere isolates compared to 32 % of bulk samples.

5.6. Discussion

A representative clone library was produced containing α and β KSs. These genes were amplified, using polyketide specific primers, from total community DNA. The site chosen had previously been shown to contain type II PKS genes using both a

direct and indirect isolation procedure. The crop site chosen was cultivated with bean plants and both rhizosphere and bulk soil samples were examined. Following PCR amplification, from total community DNA, the samples were cloned using a TA vector cloning kit (Invitrogen). Positive transformants were picked and a representative number of clones sequenced using sequencing primers (T7 and M13) specific to the vector backbone.

PKS sequences were confirmed using the BLAST database at the NCBI and partial sequences (400 bp) prepared for phylogenetic analysis. Partial sequences (400 bp) from actinomycete type strains were examined for their ability to confer phylogenetic relatedness to soil clone sequences. Phylogenetic analysis of these partial sequences showed clusters consistent with previously published phylogenetic studies (Hopwood, 1997; Seow *et al.*, 1997). Analysis of both KSα and KSβ sequences showed genes responsible for polyketide spore pigments clustering separately from those responsible for antibiotic and other bioactive polyketide compounds. High bootstrap values gave confidence to these observations. This suggests that the partial sequences chosen for this analysis can be used to determine the phylogenetic relationships between type II PKS α and β KS genes.

All soil clone sequences analysed were found to cluster within published PKS sequences. Rhizosphere samples appeared to show the greatest diversity with sequences observed to be representative of both spore pigment and antibiotic gene clusters. No sequences were isolated from total community DNA extracted from bulk soil that clustered with spore pigment polyketide clusters. This may indicate a greater distribution of type II PKS genes encoding spore pigments within the

rhizosphere when compared to bulk soil. Further sequencing will be required though to test the statistical validity of this hypothesis. Seow *et al.* (1997) isolated KSβ sequences from a forest area near Vancouver, British Columbia, Canada. Analysis of 2 sequences showed them to cluster within type II polyketide antibiotic PKS clusters. This may suggest that spore pigment PKS genes are less widely distributed when compared to PKS genes responsible for antibiotic production. This may be due to bias in the primer design preferentially amplifying type II PKS genes encoding antibiotic polyketides.

Lower diversity was observed in bulk soil samples when compared to type strain and rhizosphere clone sequences. A number of bulk soil clone sequences clustered together forming a distinct group when compared to published sequences. This trend was observed in both KSα and β sequences. Analysis of the branch lengths in this cluster group suggests some of these sequences are closely related. This may represent a novel group of KS genes within the soil community. Soil clone sequences were found to be between 80-92 % similar to published sequences (Chapter 9). Metsä-Ketelä *et al.* (1999) suggested that this level of nucleotide similarity may code for polyketide chains with significant structural diversity. The approach discussed in Chapter 6 may illustrate a way to access this diversity with implications in microbial biotechnology.

KSα and KSβ sequences were compared to each other in cases where sequences for both genes were available. In almost every case the α and β KS genes clustered together as either spore pigment or antibiotic production genes. Only one case was

observed where this was not the case. This clone isolated from the rhizosphere (R4) contained a gene clustering within the spore polyketide genes (KSα) and a gene clustering within the antibiotic polyketide genes (KSβ). This may suggest acquisition of KS genes from separate sources possibly by horizontal gene transfer. Evidence suggests that α and β KS genes are transcriptionally linked and that the gene products form a protein dimer. Heterologous expression of different KS subunits often results in low levels of polyketide chain elongation (C.R. Hutchinson, pers. comm.). This may explain why α and β KS genes appear to cluster together and may implicate horizontal gene transfer of the KS genes as a pair rather than as single units.

Actinomycete isolates were picked and 200 screened for the presence of type II PKS genes. These isolates were picked in equal numbers from rhizosphere (100) and bulk samples (100). In total 36 % of these isolates were found to contain type II PKS genes. A higher percentage of rhizosphere isolates (40 %) were found to contain type II PKS genes compared to bulk soil isolates (32 %). Sequencing of these isolates will determine if direct extraction of type II PKS genes from total community DNA when compared to an indirect isolation method yields more diverse type II PKS sequences.

Due to the significant interest in the nucleotide and amino acid sequences of genes encoding polyketide metabolites a large database has become available containing these sequences. This has allowed phylogenetic analysis to deduce evolutionary relationships and consider the origin of these biosynthetic pathways (Hopwood,

1997). Sequence comparison suggests that PKSs and FASs may well have a common evolutionary origin and a strong relatedness is seen between the condensing enzyme subunits of both type I and type II PKSs (Siggaard-Anderson, 1993). A similar relatedness can be seen with genes encoding ACPs (W.P. Revill, pers. comm.). An interesting observation is that at present no FAS or PKS sequences have been observed in Archaea. Sequencing of the *Methanococcus jannaschii* genome failed to reveal any KS genes (Bult *et al.*, 1996). This suggests possibly that the genes for FASs and PKSs evolved after Archaea became separate from Bacteria and Eucarya.

A theory may be that an early ancestor developed a primitive condensing enzyme that recruited other functions, for example ACP or AT domains, to become more efficient. This could give rise to a primitive PKS followed by acquisition of a reductive enzyme to give rise to a FAS (Hopwood, 1997). A hypothesis may be that early type II PKS clusters could join to give rise to a type I PKS system. Evidence suggested by Schwecke *et al.* (1995) showed that the 14 module rapamycin PKS shares almost identical subunits. Motamedi *et al.* (1997) showed that certain subunits within the rapamycin PKS are more similar to those in the FK506 cluster than the FK506 subunits themselves. This possibly suggests a recent acquisition of these genes by horizontal gene transfer. These include duplication of PKS genes to give rise to multiple modules. Recruitment of certain enzymes to give rise to greater or more controlled diversity, or horizontal gene transfer to acquire new genes or speed up adaptation. All these have particular relevance within the soil community. Weiner *et al.* (1998) have observed evidence of the transfer of streptomycin resistance genes in natural *Streptomyces* communities. This study implicated that

horizontal gene transfer may be an important event in the evolution of biosynthetic pathways (Weiner *et al.*, 1988).

5.6.1. Contribution to the field of study

- Analysis of type II PKS α and β KS genes from a specified site as determined by PKS specific PCR on total community DNA.
- Construction of a representative clone library containing PKS genes amplified directly from total community DNA.
- Comparison of the phylogenetic relationship between α and β KS genes isolated directly from bulk and rhizosphere samples from a single field site.
- Analysis of α and β KS gene pairs amplified from total community DNA.

5.6.2. Future Work

- Further sequencing of soil clones to determine if trends observed are representative of a larger sample number.
- Construction and analysis of clone libraries from other Cuban sites to determine if diverse sequences are site specific or representative of the Cuban actinomycete community.
- Analysis of type II PKS sequences from actinomycete isolates to determine if the diversity observed in total community DNA is greater than the culturable fraction.

Chapter 6

Development of a method for heterologous expression of near full length type II PKS α and β ketosynthase genes from environmental DNA

In collaboration with;

6.1. Introduction

The potential for producing new "hybrid" antibiotic structures by engineering novel combinations of antibiotic genes from different organisms was first demonstrated by transferring segments of the *act* biosynthetic cluster into other streptomycete strains that produce related polyketide antibiotics (Hopwood *et al.*, 1985). Hybrid compounds produced in these early experiments were polyketides but the recombinant genes used were responsible for tailoring steps in the biosynthetic pathway rather than genes affecting polyketide chain assembly or immediate post-assembly modification (Epp *et al.*, 1989). Initially it was unclear as to whether these chain assembly genes could be mixed to form functional hybrid synthases due to protein interactions allowing only certain synthase subunits to work together. Work carried out by Bartel *et al.* (1990) showed that novel polyketides could be made by "mixing and matching" heterologous subunits from different PKS clusters. Further experiments demonstrated the ability of a systematic mix-and–match approach to produce polyketide structures generated by recombinants containing hybrid PKS gene clusters (Hopwood, 1997). Polyketide genes are particularly suited to producing product variation in this way and very few other secondary metabolism enzymes can display as much plasticity in substrate utilisation (Seow *et al.*, 1997).

It is now possible with the understanding of polyketide biosynthesis to contemplate adapting PKS gene clusters to generate large libraries of unnatural natural products. Libraries produced in this way offer unique opportunities for drug discovery and development since they combine combinatorial techniques with the proven record of polyketide natural products as useful chemical agents (Tsoi and Khosla, 1995).

Using the minimal type II PKS and its associated "design rules" it is possible to produce engineered aromatic polyketides (McDaniel *et al.*, 1994). Due to this demonstration of combinatorial biosynthesis experiments have been conducted to search for novel PKS-encoding genes including those not producing pharmacologically useful products. The highly conserved nature of PKS genes allows them to be used to screen DNA libraries. Ruan *et al.* (1997) used this approach to isolate a second PKS cluster from a rapamycin producing strain of *S. hygroscopicus*. Another target is fungal symbionts of lichens which are well known for both their polyketide products and notoriously slow growth making laboratory growth impractical (Bentley and Bennett, 1999).

New approaches to drug discovery are now being developed including isolation of novel biosynthetic genes from total community DNA without the need for initial isolation. This allows rapid isolation of DNA from environmental samples without the need for isolation and cultivation avoiding the inherent bias present in these methods. This demonstrates a potential way to gain access to a more extensive range of microbial diversity and to biosynthetic pathways whose products can be tested for biological applications. Access to this pool of biosynthetic genes from unculturable microorganisms has significant implications for microbial biotechnology (Seow *et al.*, 1997). This approach is now being employed by a number of pharmaceutical companies including TerraGen Diversity Inc. (Vancouver, Canada), MaxyGen Inc. (Redwood City, California, USA) and Diversa Corp. (San Diego, California, USA) on the basis that it will ideally lead to more compounds being presented for screening for potential biological activity.

Metsä-Ketalä *et al.* (1999) designed degenerate oligonucleotide primers to amplify KSα genes from bacterial isolates. This method was designed as a rapid method for pharmaceutical companies to screen strain collections. Phylogenetic analysis of sequences isolated using this method were found to align with both antibiotic producing and spore pigment gene clusters. The limitation of this system is that the primer pair used only amplifies a small part of the KSα so any amplified gene products will only be of use for phylogenetic analysis and screening isolates.

Seow *et al.* (1997) used PCR primers derived from conserved regions of known ketosynthase (KSα) and ACP genes to amplify KSβ genes from 2 unclassified streptomycetes and 2 clones derived from total community DNA. Sequence analysis and subsequent phylogenetic analysis deduced these gene sequences aligned to antibiotic producing gene clusters. Initial evidence suggested that these genes may be responsible when coupled with the *actI*-Orf1 KSα for the production of octaketides. This could be due to a number of reasons including bias in the PCR protocol due to degeneracy in the primers or the *actI*-Orf1 KSα having more influence over chain length than initially believed. The yield from the recombinant systems was very low possibly due to incompatibility in heterologous KS protein interactions. This provided difficulties in determining the structure of the produced hybrid polyketides (C.R. Hutchinson, pers. comm.).

The development of a method to allow amplification of KSαβ gene pairs will allow subsequent expression without the inherent problems associated with heterologous protein interactions between 2 different synthase subunits. Subsequent sequencing

of any amplified products would also allow phylogenetic analysis of the polyketide genes to be undertaken.

6.1.1. Aims

- To develop a system to amplify near full length α and β ketosynthase genes from type II PKS gene clusters.
- To use this method to amplify PKS genes from type strain DNA and total community DNA.
- To develop an expression system that will allow the expression of a minimal PKS consisting of PCR amplified KS genes and *tcmM*.

6.2. Development of a PCR method to amplify near full length α and β ketosynthase genes

PCR primers were developed to amplify near full length α and β ketosynthase genes from type II PKS gene clusters. Degenerate oligonucleotides were designed from amino acid pileup sequences of 12 type II PKS gene clusters representative of both antibiotic producing and spore pigment biosynthetic clusters (Figure 6.1 and 6.2). This primer pair was designed to amplify a region of approximately 2.5 kb from type II PKS gene clusters spanning near full length α and β ketosynthase genes (Table 6.1).

a) Peptide translation at the N-terminal region

Act	- - - - - MKRRV	VITGVGVRAP	GGNGTRQFWE
Dau	- - - - - MNRRI	VITGIGVVAP	A - VGTKPFWE
Fren	- - - - - MNRQV	AVTGIGVVAP	GGIGRKPYWE
Gra	- - - - - MTRRV	VITGVGVRAP	GGSGTKEFWD
Gris	- - - - - MERRA	VITGIGVCAP	CGTGVDSFWE
Jad	- - - - MTARRV	VITGIEVLAP	GGTGSKAFWN
Mtm	- - - - - MNRRV	VITGIGVVAP	GAVGVEAFWE
Otc	MSKIHDARRV	VITGIGVVAP	GDVGTKPFWE
Sno	- MKESINRRV	VITGIGIVAP	DATGVKPFWD
Tcm	-MTRHAEKRV	VITGIGVRAP	GGAGTAAFWD
Urd	-MSGAHSRRV	VITGIGVTAP	GGVGSKNFWS
WhiE	- - - - - MTRRRV	AVTGIGVVAP	GGIGTPQFWR
Consensus	-MK - -MNRRV	VITGIGVVAP	GGVGTKPFWE

b) Nucleotide sequence at the N-terminal region

Act	CGC	AGA	GTC	GTC	ATC	ACG	GGC	GTC	GG
Dau	CGG	CGG	ATC	GTC	ATC	ACC	GGG	ATC	GG
Fren	CGA	CAA	GTC	GCC	GTC	ACC	GGC	ATC	GG
Gra	CGA	CGC	GTA	GTG	ATC	ACC	GGG	GTC	GG
Gris	CGC	CGG	GCA	GTG	ATC	ACG	GGA	ATC	GG
Jad	CGA	CGC	GTC	GTC	ATC	ACC	GGC	ATC	GA
Mtm	CGT	CGC	GTC	GTC	ATC	ACC	GGC	ATC	GG
Otc	CGA	CGC	GTC	GTA	ATC	ACA	GGG	ATC	GG
Sno	CGT	CGC	GTG	GTC	ATC	ACC	GGA	ATA	GG
Tcm	AAG	CGG	GTG	GTG	ATC	ACC	GGC	ATC	GG
Urd	CGA	CGC	GTC	GTG	ATC	ACC	GGC	ATC	GG
WhiE	CGC	CGG	GTC	GCC	GTC	ACG	GGC	ATA	GG
Consensus	CGA	CGC	GTC	GTC	ATC	ACC	GGC	ATC	GG

c) Consensus primer design showing peptide translation

Peptide		R	R	V	V	I	T	G	I
Primer	CGN	CGS	GTC	GTC	ATC	ACC	GGC	ATC	GG

Figure 6.1. **Design of forward PCR primer.** For key see Table 6.1.

a) Peptide translation at the C-terminal region

Act	PRTALVLARG	RWGFNSAAVL	- - RRFAPTP	
Dau	ARAALLVARG	TGGFNSALVV	RGAA	
Fren	LRTALVLARG	HGGFNAAVVV	RGRRRPRTA	
Gra	LSAALVLARG	RHGFNSAVVV	TLRGSDHRRP	T
Gris	VGTALVLARG	HGGFNSAVVL	RAVD	
Jad	VNTALVIARG	HGGFNSAMVV	RSAN	
Mtm	LRAALVLARG	RGGFNAAAVV	RALS	
Otc	LRHALVLARG	HGGFNSAMVV	SGRD	
Sno	LARALVLARG	RGGFNAAMVV	AGPRAETR	
Tcm	VDTALVVARG	MGGFNSALVV	RRHG	
Urd	VRSALVIARG	HGGFNSAVVV	RAVG	
WhiE	PRTALVLARG	LMGSNSALVL	RRGAVPPEGR	
Consensus	LRTALVLARG	HGGFNSA - VV	RGR - - - P	

b) Nucleotide sequence at the C-terminal region

Act	GCC AGG	GGC	CGC	TGG	GGC	TTC	AAC	TCG	GCG
Dau	GCC CGC	GGC	TAC	GGG	GGC	TTC	AAC	AGC	GCT
Fren	GCGCGC	GGC	CAC	GGC	GGC	TTC	AAC	GCG	GCC
Gra	GCG CGC	GGC	CGC	CAC	GGC	TTC	AAC	TCC	GCG
Gris	GCGCGA	GGG	CAC	GGC	GGA	TTC	AAC	TCG	GCC
Jad	GCCCGC	GGT	CAC	GGC	GGC	TTC	AAC	TCC	GCG
Mtm	GGCTCG	GCG	GTG	GTC	GTG	TCC	AGC	AGA	CGC
Otc	GCCCGC	GGT	CAC	GGC	GGG	TTC	AAC	TCC	GCG
Sno	GCGCGG	GGC	CGG	GGC	GGG	TTC	AAT	GCG	GCG
Tcm	GCGCGG	GGC	ATG	GGC	GGA	TTC	AAC	TCC	GCC
Urd	GCC CGC	GGC	CAC	GGC	GGT	TTC	AAC	TCC	GCG
WhiE	GCGCGC	GGC	CTC	ATG	GGG	TCG	AAC	TCG	GCG

Consensus	GC -	CGC	GGC	CAC	GGC	GGC	TTC	AAC	TCC
GCG									

c) Consensus primer design showing peptide translation

Peptide	A	R	G	H	G	G	F	N	S
Primer	GCS	CGC	GGC	CAC	GGC	GGS	TTC	AAC	TCS
	A								
	GCG								

Figure 6.2. **Design of reverse PCR primer.** For key see Table 6.1

Strain	PKS Product	Protein	Product size
S. coelicolor	actinorhodin	Act	2497 bp
S. peucetius	daunorubicin	Dau	2542 bp
S. roseofulvus	frenolicin	Fren	2638 bp
S. violaceoruber	granaticin	Gra	2509 bp
S. griseus	griseusin	Gris	2503 bp
S. venezuelae	jadomycin	Jad	2479 bp
S. argillaceus	mithramycin	Mtm	2491 bp
S. rimosus	oxytetracyclin	Otc	
S. nogalater	nogalamycin	Sno	2518 bp
S. glaucescens	tetracenomycin	Tcm	2506 bp
S. fradiae	urdamycin	Urd	2503 bp
S. coelicolor	spore pigment	Whi	2542 bp

Table 6.1. PKS genes used for the PCR primer design and the predicted size of the PCR generated product.

The position of the forward primer was chosen due to high levels of peptide sequence similarity at a position closest to the 5' end of the KSα. No high level of similarity between type strains is seen in the region proceeding the 5' start position of the KSα. The reverse primer site was chosen within the KSβ to ensure that any observations seen during expression studies were not due to remnants of ACP nucleotides present in PCR products. A *Pst*I restriction site was incorporated into the forward primer and an *Mro*I restriction site into the reverse primer. This was to facilitate downstream processing of amplified PCR products. A ribosome binding site (RBS) and start / stop codons were also incorporated to insure fidelity in expression studies (Figure 6.3).

6.3. Testing primers against type strains

DNA was extracted from *S. glaucescens* and *S. peucetius* using a small scale DNA extraction method (2.11) and diluted to 100 ng μl^{-1}. PCR products were produced using a standard PCR protocol (2.13) using Vent polymerase to give blunt ended PCR products (Figure 6.4).

Forward primer :

5' -C CTG CAG *GGA GGA* TGC GNC GSG TCG TCA TCA
***Pst*I** **RBS** **start**

CCG GCA TCG G – 3'

Reverse primer :

5' -C TCC GGA *TCA* CGC SGA GTT GAA SCC GCC GTG
***Mro*I** **stop**

GCC GCG SGC – 3'

Figure 6.3. Final primer design showing restriction sites, the ribosome binding site and transcriptional start / stop sites.

6.4. Construction of PKS expression vectors and cloning strategy

PCR amplified products were gel extracted, phosphorylated and purified using phenol chloroform precipitation (2.13.1). Ligation and transformations were carried out as described in the materials and methods (2.14).

6.5. pUC18

pUC18 (Amersham Pharmacia) was used as the basic vector backbone for assembling the expression vectors and cloning PCR products. This was purchased

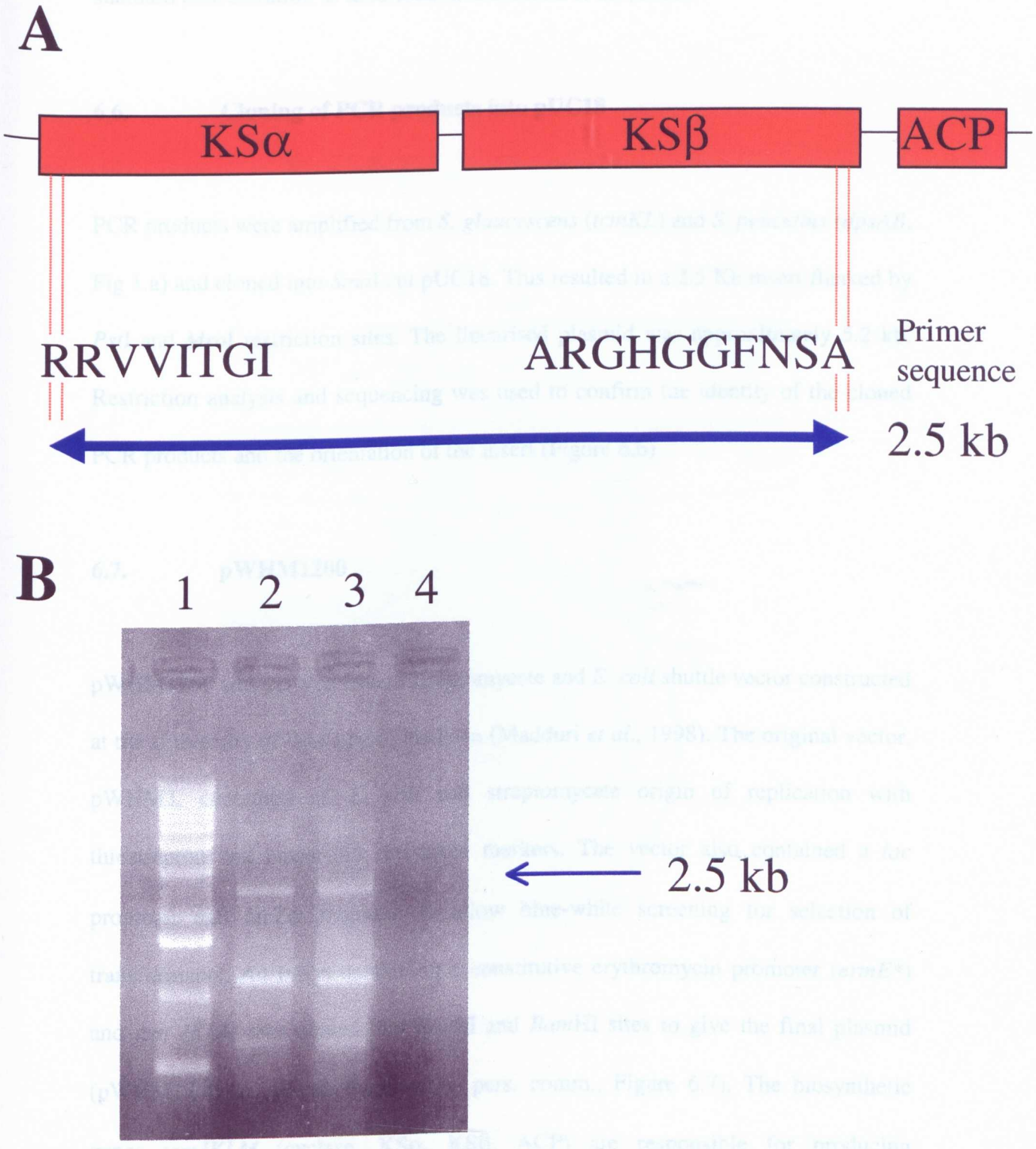


Figure 6.4. (A) Primer set for the amplification of near full length α and β ketosynthase genes. (B) Amplification of α and β ketosynthase genes from streptomycete type strains. Lane 1. 1 kb marker, 2. *S. glaucescens* (*tcm*), 3. *S. peucetius* (*dps*), 4. ddH_20.

as a *Sma*I cut, dephosphorylated linear plasmid (2686 bp, figure 6.5) and used at a standard concentration as described in Sambrook *et al.* (1989).

6.6. Cloning of PCR products into pUC18

PCR products were amplified from *S. glaucescens* (*tcmKL*) and *S. peucetius* (*dpsAB*, Fig 1.a) and cloned into *Sma*I cut pUC18. This resulted in a 2.5 Kb insert flanked by *Pst*I and *Mro*I restriction sites. The linearised plasmid was approximately 5.2 kb. Restriction analysis and sequencing was used to confirm the identity of the cloned PCR products and the orientation of the insert (Figure 6.6).

6.7. pWHM1200

pWHM1200 was derived from a streptomycete and *E. coli* shuttle vector constructed at the University of Wisconsin, Madison (Madduri *et al.*, 1998). The original vector, pWHM3, contained an *E. coli* and streptomycete origin of replication with thiostrepton and ampicillin resistance markers. The vector also contained a *lac* promoter and *lacZ*α fragment to allow blue-white screening for selection of transformants. An insert containing a constitutive erythromycin promoter (*ermE**) and *tcm JKLM* was cloned into *Eco*RI and *Bam*HI sites to give the final plasmid (pWHM1200; E. Wendt-Pienkowski, pers. comm., Figure 6.7). The biosynthetic genes *tcmJKLM* (cyclase, KSα, KSβ, ACP) are responsible for producing tetracenomycin F2 a precursor of the final polyketide, tetracenomycin C (Hutchinson *et al.*, 1993). The *ermE** promoter was constructed by M.J. Bibb to be a strong, constitutive promoter. The *ermE* promotor of the *ermE* gene of

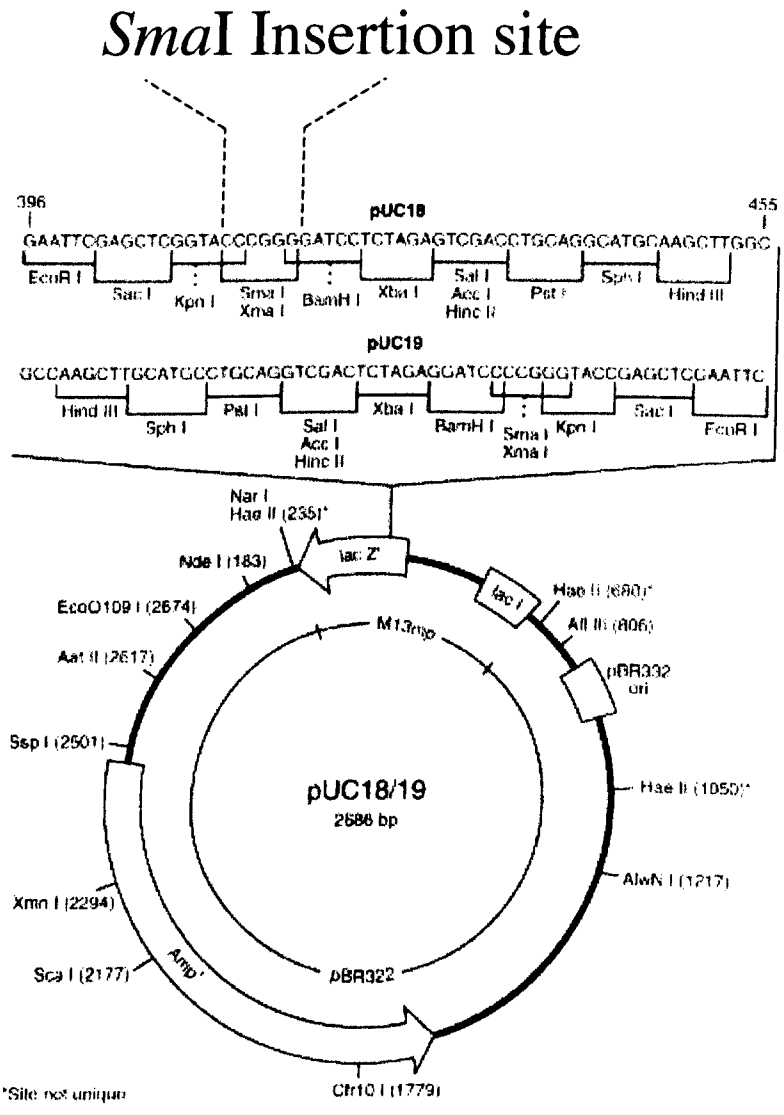


Figure 6.5. Plasmid map of pUC18 showing incorporated restriction sites. Also shown are genes for antibiotic resistance and selection of transformants. The *Sma*I restriction site is indicated as the insertion site for cloning used in this study.

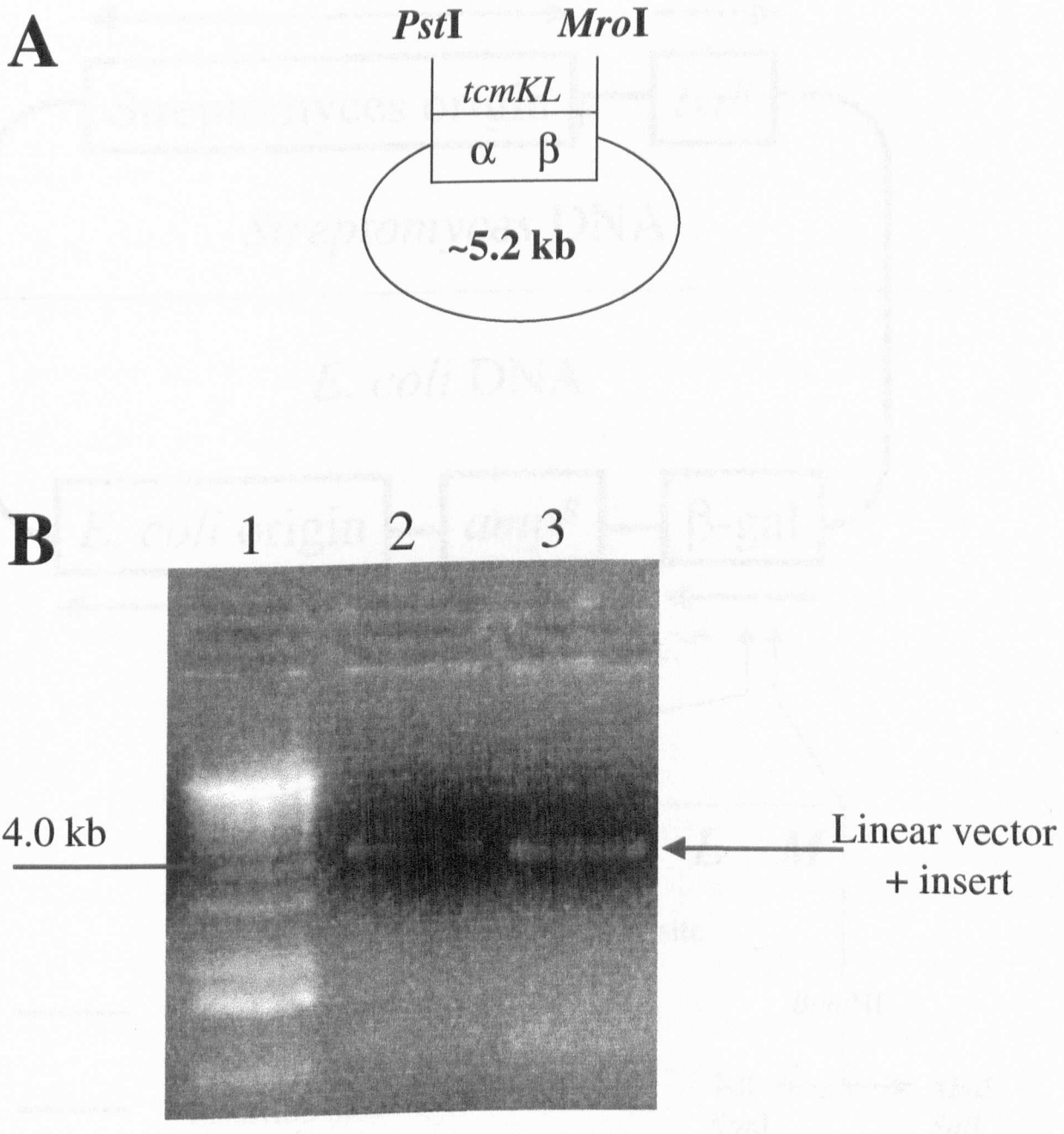


Figure 6.6. Cloning and restriction analysis of the PCR product containing the α and β ketosynthase (*tcmKL*) from *S. glaucescens*. **(A)** A 2.5 kb fragment was amplified from *S. glaucescens* and ligated into pUC18. (B) The resulting plasmid was digested with *Pst*I and *Mro*I to ensure the linear product was the correct size and that the predicted restriction sites incorporated from the PCR protocol were present. Lane 1. 1 kb marker, 2. *Pst*I cut (5163, 23 bp), 3. *Mro*I cut (5186 bp). The insert was also confirmed by sequencing.

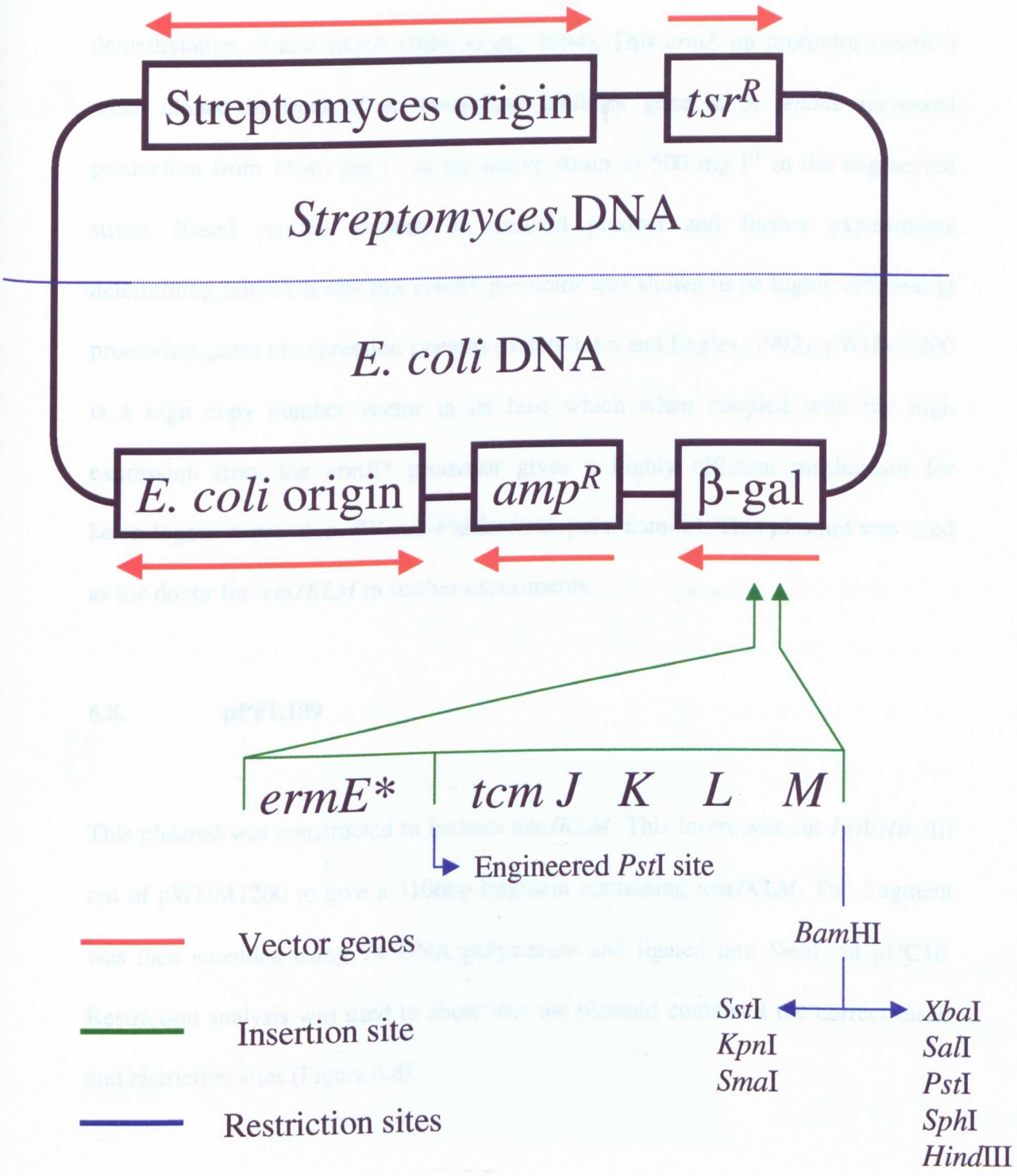


Figure 6.7. Plasmid map, insert and restriction sites of pWHM1200. The insert (*ermE** and *tcmJKLM*) was cloned into *Eco*RI and *Bam*HI restriction sites (Wendt-Pienkowski, pers. comm.).

Saccharopolyspora erythraea confers resistance to erythromycin by N^6-demethylation of 23S rRNA (Bibb *et al.*, 1994). This *ermE* up promotor (*ermE**) when cloned in front of an α-amylase inhibitor gene in *S. tendae* increased production from 15-40 mg l^{-1} in the native strain to 500 mg l^{-1} in the engineered strain. Based on the amount of secreted product and further experiments determining mRNA levels this *ermE** promotor was shown to be highly efficient at promoting genes in expression systems (Schitt-John and Engles, 1992). pWHM1200 is a high copy number vector in its host which when coupled with the high expression from the *ermE** promotor gives a highly efficient mechanism for heterologous expression. (Wendt-Pienkowski, pers. comm.). This plasmid was used as the donor for *tcmJKLM* in further experiments.

6.8. pPFL109

This plasmid was constructed to include *tcmJKLM*. This insert was cut *PstI/HindIII* out of pWHM1200 to give a 3106bp fragment containing *tcmJKLM*. The fragment was then amended using T4 DNA polymerase and ligated into *Sma*I cut pUC18. Restriction analysis was used to show that the plasmid contained the correct insert and restriction sites (Figure 6.8).

6.9. Addition of *tcmM* to cloned PCR products.

pPFL109 was cut using *Pst*I and *Mro*I to give 2 linear products. The vector backbone containing pUC18 and *tcmM* (2876 bp) and *tcm JKL* (2916 bp). Cloned PCR products were cut using *Pst*I and *Mro*I to give 2 linear products. The α and β

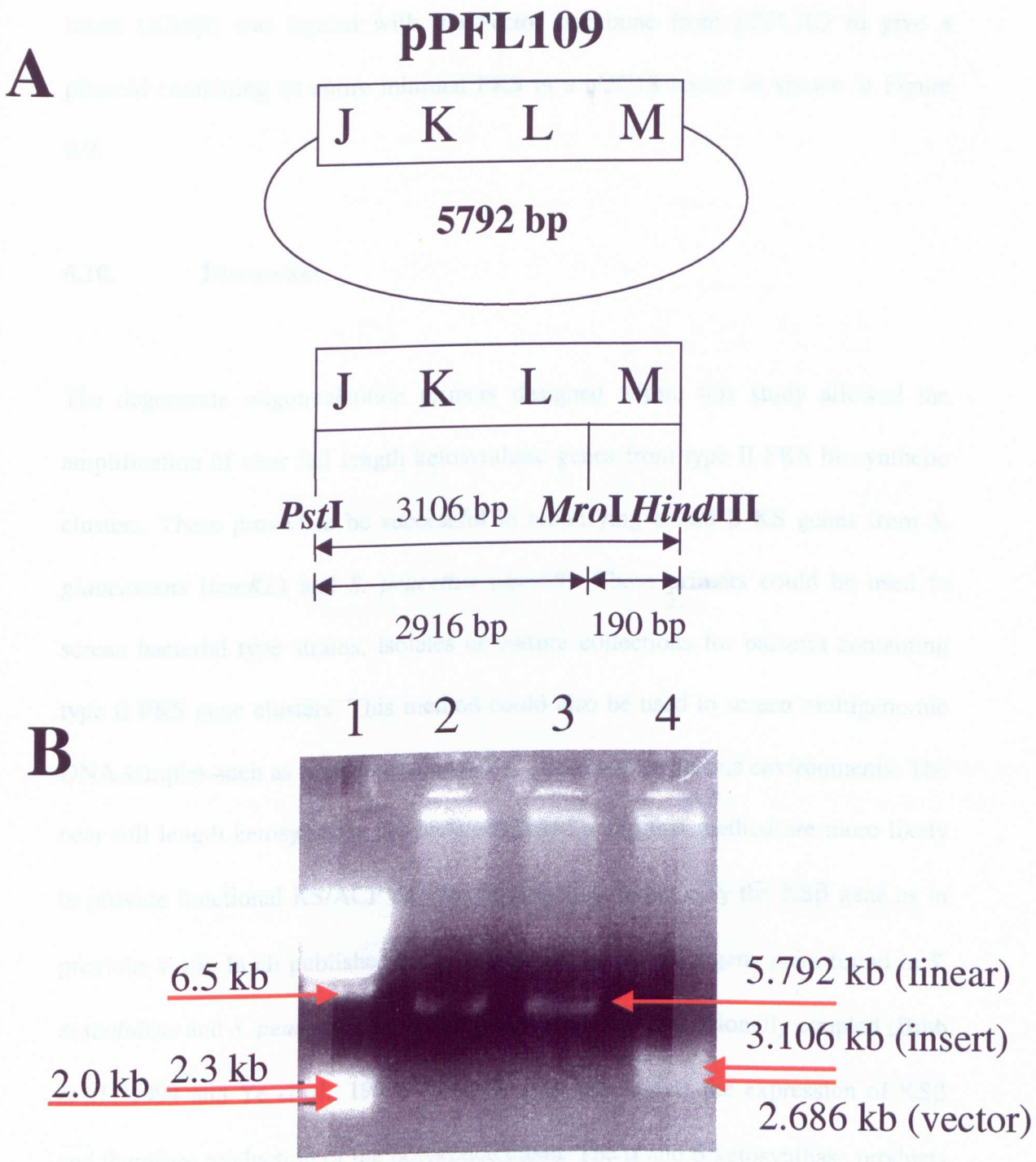


Figure 6.8. Plasmid map of pPFL109. **(A)** A 3106 bp fragment was cut *Pst*I / *Hind*III out of pWHM1200 and ligated into *Sma*I cut pUC18 to give pPFL109 containing *tcmJKLM*. **(B)** This was confirmed by restriction analysis on the resulting plasmid. Lane 1. *Hind*III cut λ, 2. *Pst*I cut, 3. *Hind*III cut, 4. *Pst*I / *Hind*III cut.

KS (~2.5 kb) and the pUC18 vector backbone (2696 bp). After size selection the insert (KSαβ) was ligated with the vector backbone from pPFL109 to give a plasmid containing an entire minimal PKS in a pUC18 vector as shown in Figure 6.9.

6.10. Discussion

The degenerate oligonucleotide primers designed within this study allowed the amplification of near full length ketosynthase genes from type II PKS biosynthetic clusters. These proved to be successful in amplifying α and β KS genes from *S. glaucescens* (*tcmKL*) and *S. peucetius* (*dpsAB*). These primers could be used to screen bacterial type strains, isolates or culture collections for bacteria containing type II PKS gene clusters. This method could also be used to screen multigenomic DNA samples such as total community DNA from soil or marine environments. The near full length ketosynthase products amplified using this method are more likely to provide functional KS/ACP enzyme hybrids than using only the KSβ gene as in previous work. In all published cases, except the KSα / KSβ gene pairs found in *S. roseofulvus* and *S. peucetius*, the α and β KS genes are translationally coupled (Bibb *et al.*, 1994 and Ye *et al.*, 1994). Disruption of this affects the expression of KSβ and therefore production of the polyketide chain. The α and β ketosynthase products are also believed to combine to form a dimer. Co-expression of single KS genes from different PKS pathways is believed to give inefficient polyketide chain elongation due to incompatibility in heterologous protein interactions (C.R. Hutchinson, pers. comm.).

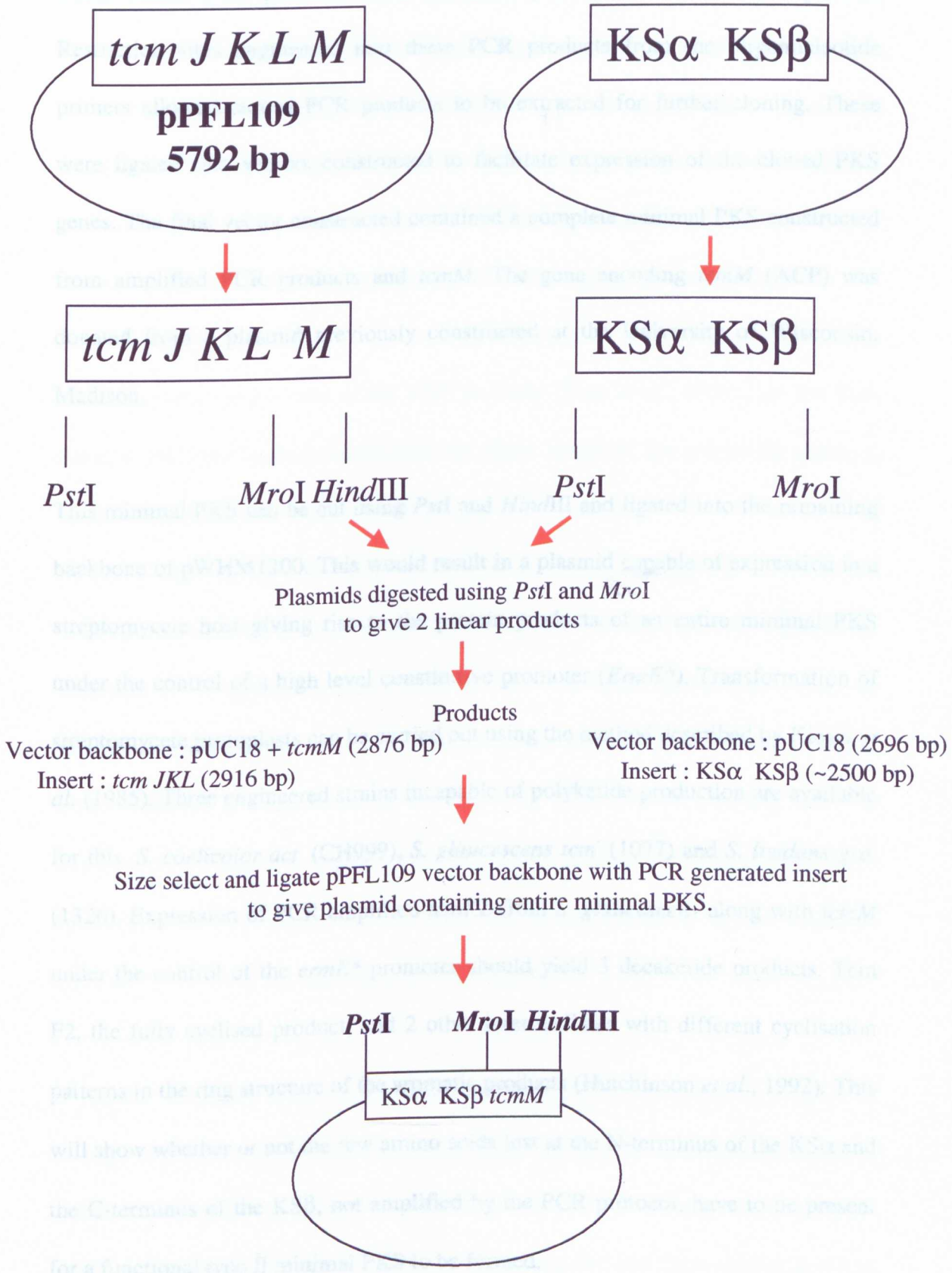


Figure 6.9. Addition of an ACP (*tcmM*) to cloned α and β ketosynthase genes to give a entire minimal PKS.

Blunt ended PCR products were successfully cloned into *SmaI* cut pUC18. Restriction sites engineered into these PCR products from the oligonucleotide primers allowed cloned PCR products to be extracted for further cloning. These were ligated into vectors constructed to facilitate expression of the cloned PKS genes. The final vector constructed contained a complete minimal PKS constructed from amplified PCR products and *tcmM*. The gene encoding *tcmM* (ACP) was donated from a plasmid previously constructed at the University of Wisconsin, Madison.

This minimal PKS can be cut using *Pst*I and *Hind*III and ligated into the remaining backbone of pWHM1200. This would result in a plasmid capable of expression in a streptomycete host giving rise to the protein products of an entire minimal PKS under the control of a high level constitutive promoter (*ErmE**). Transformation of streptomycete protoplasts can be carried out using the method described by Kieser *et al.* (1985). Three engineered strains incapable of polyketide production are available for this, *S. coelicolor act*$^{-}$ (CH999), *S. glaucescens tcm*$^{-}$ (1077) and *S. lividans gra*$^{-}$ (1326). Expression of PCR amplified *tcmKL* from *S. glaucescens* along with *tcmM* under the control of the *ermE** promoter should yield 3 decaketide products. Tcm F2, the fully cyclised product, and 2 other intermediates with different cyclisation patterns in the ring structure of the aromatic products (Hutchinson *et al.*, 1992). This will show whether or not the few amino acids lost at the N-terminus of the KSα and the C-terminus of the KSβ, not amplified by the PCR protocol, have to be present for a functional type II minimal PKS to be formed.

Further cloning experiments could be carried out to add cyclase genes to the expression system. Addition of *tcmJ* to give expression of *tcmJKLM* should increase the concentration of Tcm F2 produced. Tcm F2 is the natural precursor to Tcm F1 in the biosynthetic pathway of Tcm C. Intermediates produced with different cyclisation patterns are caused by a lack of fidelity in the biosynthetic pathway if cyclisation genes, including *tcmJ*, are absent (P.F. Long, pers. comm.).

It is believed as in the case of the *whiE* pathway (Shen *et al.*, 1999) that the Tcm minimal PKS has limited control over the chain length of the polyketide chain. A range of products will be produced and all will be based on the condensation of acetate units. The hypothesis is that the Tcm ACP can be acylated by a range of carboxylic acids but it is the KSα or KSβ (or KSαβ) that show substrate specificity. Therefore by swapping the KSα / KSβ of Tcm with other homologues it may be possible to isolate products with a range of chain lengths based on acetate starter units. A similar strategy based on the Dps PKS from *S. peucetius* may result in similar products based on a propionate starter unit. Adding further cyclase genes to these minimal PKS genes should result in further variation in the observed cyclisation pattern of the fused ring aromatic products.

A genetic approach based on bacterial aromatic PKSs has led to a number of insights into molecular recognition, chain length specificity and regiospecificity in carbon chain cyclisation. The biosynthetic pathways of PKSs show a remarkable ability for controlled biosynthesis of complex biomolecules from simple building blocks. The lessons learnt in genetic and protein engineering of PKS structures and mechanisms is likely to lead to many ideas that could be exploited for the directed

synthesis of novel hybrid compounds with biological and pharmaceutical activity (Khosla *et al.*, 1999).

6.10.1. Contribution to the field of study

- A set of degenerate oligonucleotide primers were designed enabling the amplification of near full length α and β ketosynthase genes from streptomycete type strains.
- A cloning strategy was employed to allow these genes to be expressed in combination with ACP to give a functional minimal PKS.
- The described system was used to clone the α and β ketosynthases from the *tcm* (*tcmKL*) system of *S. glaucescens* with *tcmM* (ACP) to use as a control for checking hybrid products.

6.10.2. Future Work

- Application of the degenerate oligonucleotide primers to total community DNA samples to produce clone libraries of environmental α and β ketosynthase genes.
- Isolate phylogenetically diverse α and β ketosynthase genes from clone libraries using restriction analysis and representative sequencing of diverse clones
- Clone and express identified diverse sequences using described expression vectors.
- Chemical characterisation of new hybrid compounds using HPLC and NMR.

Chapter 7

General Discussion

7.1. General Discussion

Molecular studies have been applied to terrestrial systems to study species abundance and to examine community diversity (Stackebrandt, 1992). Studies include those on methanotrophs (McDonald and Murrell, 1997) and microorganisms capable of metabolising heavy metals (Hart *et al.*, 1998). The use of this type of approach to study antibiotic genes has not been widely addressed. Huddleston *et al.* (1997) and Weiner *et al.* (1989) have examined the distribution of streptomycin producing *Streptomyces* spp. to look for evidence of gene transfer within the natural environment. Seow *et al.* (1997) extracted type II PKS genes from total community DNA but did not address ecological questions. This study has analysed the distribution and diversity of type II PKS genes in total community DNA and actinomycete isolates using a combination of molecular ecology techniques.

Molecular techniques were used to study the distribution of type II PKS genes in total community DNA. Oligonucleotide PCR primers were designed to amplify a region spanning 2 β-ketoacylsynthase (KS) genes. These genes, KSα and KSβ, have been observed in every type II PKS cluster described in the literature and are usually found to be immediately upstream of an acyl carrier protein. These 3 genes form a conserved region which is responsible for a number of key events in polyketide biosynthesis including starter unit specificity and first ring cyclisation (Hopwood, 1997).

This PCR based detection method was used to screen total community DNA extracted from a number of diverse soils freshly sampled from Cuba. Lynch and Painting (1980) observed greater microbial diversity in tropical rhizosphere soil. It is believed that tropical soil favours antibiosis due to elevated temperature and nutrient supply. Standage (1998) studied the distribution of the gentamicin resistance methyltransferase gene (*grm*) in soil communities. The *grm* gene was observed to be more widely distributed in Mediterranean and sub-tropical soils. Previous studies had indicated that certain actinomycete species are more prevalent in humid soils (Lechevalier, 1961). Temperature is believed to be an important factor in bacterial distribution and increased temperatures may allow an increase in soil biomass (Arias *et al.*, 1998).

Cuban soils were placed in 3 groups: Cuban crop soils, Botanical garden and Cayo Blanco. The Botanical garden site is a well-established facility located in Havana. Cayo Blanco is a pristine island located off the North coast of Cuba. This uninhabited island has 3 distinct ecosystems, beach, shrub and fir. Type II PKS genes were detected in 43 % of soil communities tested. Distribution was found to be dependent on the soil group. Botanical garden samples were observed to show the greatest distribution of type II PKS genes with type II PKS genes detected in 66 % of samples.

Detection of type II PKS genes in total community DNA extracted from crop soils did not correlate with isolation studies. Type II PKS genes were amplified from 44 % of Cuban crop soils tested. It was possible to isolate bacteria containing type II PKS

genes from every Cuban crop site. Standage (1998) observed similar results when screening soils for *grm* genes. This suggests that soil screening should not be used exclusively as a predictive method to determine whether or not a soil is studied further. Conversely genes have been observed in 16S rDNA clone libraries and functional gene libraries that are not representative of the culturable fraction of the community (Ward *et al.*, 1995; Hart *et al.*, 1998). This suggests that to get a true representation of the microbial community a polyphasic approach should be undertaken. This would involve comparing a number of different approaches to achieve a more representative view of the community. Although direct molecular methods avoid the problems of culturability (Liesack *et al.*, 1997) this lack of correlation between methods suggests that no one method can be used to provide a definitive view of the microbial community.

It was not possible to amplify type II PKS genes from any total community DNA samples extracted from Cayo Blanco soil. Experiments appeared to show that this was due to compounds co-extracted with total community DNA inhibiting subsequent PCR reactions for example PCR using *Taq* polymerase (Tebbe *et al.*, 1993). Although direct extraction of DNA from soil yields large amounts of DNA it can often contain impurities (Steffan *et al.*, 1998). Total community DNA extracted from Cayo Blanco soils was purified using Sephadex G50 and Chelex 100 columns (Straub *et al.*, 1994) Even after further purification using gel extraction type II PKS genes could not be amplified. It was possible though to amplify 16S rDNA genes from these samples suggesting that some PCR methods are less robust than others. Actinomycete isolates

(90 %) were found to contain type II PKS genes. These isolates were present at levels above the detection limit of the PCR. This indicates once again how reliance on a single technique could result in an underestimate of diversity.

Selective isolation on Cuban crop soils yielded a number of isolates containing type II PKS genes. Phylogenetic analysis of these isolates based on 16S rDNA and *recA* sequence showed them clustering within the streptomycetes. In agreement with published data (Eisen, 1995) trees constructed using 16S rDNA and *recA* sequence data appear highly congruent and appear to have similar resolving power. This suggests that *recA* sequences are useful for molecular systematics and may prove to be a good housekeeping gene to use along with 16S rDNAs sequences to infer phylogenetic relationships. Isolates formed their own group within published streptomycete sequences suggesting they may belong to a novel group within the soil community. Further *recA* sequencing is required though due to the lack of available actinomycete sequences in the nucleotide database.

Partial sequencing of type II PKS KSα genes from these isolates revealed similarity to published antibiotic production and spore pigment sequences. Metsä-Ketelä *et al.* (1999) observed that amino acid similarities at the levels seen in these Cuban isolates are likely to code for polyketides with novel structural variations. It was observed that chemical characterisation of polyketide compounds produced by unknown soil bacteria correlated with phylogenetic analysis. An isolate found to produce a compound similar to ε-rhodomycinone, the precursor to daunorubicin, was found to

cluster with *S. peucitius*, the producing organism. A strain that did not cluster with any published sequences was found to produce an aglycone that differed from the products of the type strains used in the study (Metsä-Ketelä *et al.*, 1999). This suggests that phylogenetic analysis of type II PKS gene sequences could be used as an initial screen to look for potentially novel biologically active products.

Type II PKS genes were amplified directly from total community DNA and cloned. Studies have shown that genes cloned directly from soil may contain considerable diversity when compared to homologous genes from isolates. McDonald and Murrell (1997) reported the isolation of novel methanol dehydrogenase genes, *mxaF*, from total community DNA. These genes were found to be distinct from *mxaF* genes isolated from cultured organisms. Type II PKS sequences amplified from total community DNA and cloned were found to show considerable diversity when compared to published sequences. Both rhizosphere and bulk soil samples were available and PCR products were cloned from each. Rhizosphere clones were found to be similar to both spore pigment and antibiotic production biosynthetic clusters. Bulk soil sequences were found to only be similar to sequences coding for polyketide antibiotics. Seow *et al.*, (1997) observed similar results in KSβ genes cloned from Vancouver soils. Expression of these KSβ sequences produced little structural diversity and most polyketides produced appeared like octaketides. There are a number of possible reasons for this. One hypothesis is that a large proportion of bacteria present in the soil community that contain type II PKS gene clusters are similar to those that produce octaketides. Alternatively this may be due to some bias

during PCR amplification despite the degeneracy built into the PCR primers. A third possibility is that the KSα (*actI* or *tcmK*) gene, co-expressed with the environmental KSβ gene, may have more control over chain length than previously thought.

Previous studies using PCR primers to amplify type II PKS genes have only amplified a part of the minimal PKS. Primers designed by Seow *et al.* (1997) amplified the whole KSβ gene but this proved unsuitable for polyketide expression studies. This was believed to be due to incompatibility between heterologous KSα and KSβ gene pairs. Primers designed by Metsä-Ketelä *et al.* (1999) only amplified a small part of the KSα gene. Primers used initially in this study spanned both the KSα and KSβ genes but did not include the whole transcript and would be unsuitable for expression studies. Primers were designed in collaboration with the University of Wisconsin, Madison to amplify near full length KSα and KSβ genes. These primers were used to successfully amplify KSα and KSβ genes from actinomycete type strains. Subsequently they have been used to amplify KSα and KSβ genes from total community DNA supplied by Terragen Inc. (P.F. Long, pers. comm.) Starting with existing constructs, vectors were constructed to successfully amplify KSα and KSβ genes along with a defined ACP (*tcmM*). These expression vectors have been used to express KSα and KSβ genes from *S. glaucescens* (*tcmJK*) to confirm the fidelity of the expression system. The cylase, *tcmJ*, can also be added to give a fully aromatic polyketide structure (P.F. Long, pers. comm.). Further work using this system with PCR products amplified from total community DNA will ideally identify KS gene pairs capable of coding for unusual polyketide structures.

This approach for isolating biosynthetic genes directly from total community DNA is being developed by a number of biotechnology companies. Although evidence suggests that screening solely total community DNA sample may miss diversity in certain samples it has advantages in terms of speed over traditional culturing methods. Terragen Inc. (Vancouver, Canada; Davies, 1997) is interested in amplifying novel polyketide and other biosynthetic genes from the natural environment. Community DNA is extracted from the sample using a number of different methods and then pooled. Although this may reduce bias due to unsuccessful cell lysis under certain conditions there is a danger that strongly represented sequences may be present in such concentrations as to be detected preferentially (C.R. Hutchinson, pers. comm.)

Today the understanding of polyketide biosynthesis has reached a point where the systems can be adapted to suit our own means. Generating libraries of unnatural natural products is now possible with the opportunity to produce libraries combining the power of combinatorial chemistry with the proven track record of natural products (Tsoi and Khosla, 1995). In today's golden age hybrid antibiotics can be produced in the hope of finding new and useful polyketide products (Bentley and Bennett, 1999). Coupling these gene shuffling techniques techniques with molecular ecology methods is a way to access the immense pool of biosynthetic genes from unculturable microorganisms. This has significant implications for the biotechnology industry (Seow *et al.*, 1997).

Chapter 8

References

References

Amann, R.I., Ludwig, W. and Schleifer, K.-H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews.* **59**, 143-169.

Arias, C.R., Pujalte, M.J., Garay, E. and Aznar, R. (1998). Genetic relatedness among environmental, clinical and diseased-eel *Vibrio vulnificus* isolates from different regions by ribotyping and randomly amplified polymorphic DNA PCR. *Appl Environ. Microbiol.* **64**, 3403-3410.

August, P.R., Tang, L., Yoon, Y.J., Ning, S., Miller, R., Yu, T.W., Taylor, M., Hoffman, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and Floss, H.G. (1998). Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5**, 69-79.

Baltz, R.H. (1986). Mutagenesis in *Streptomycete* spp. In *Manual of Industrial Microbiology and Biotechnology*, pp. 184-190. Edited by A.L. Demain and N.A. Solomon. American Society for Microbiology.

Bao, W., Sheldon, P.J., Wendt-Pienkowski, E. and Hutchinson, C.R. (1999). The *Streptomyces peucetius dpsC* gene determines the choice of starter unit in biosynthesis in the daunorubicin polyketide. *J. Bact.* **181,** 4690-4695.

Barns, S.M., Fundyga, R.E., Jeffries, M.W. and Pace, N.R. (1994). Remarkable archaeal diversity detected in a Yellow-Stone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA* **54**, 1609-1613.

Bartel, P.L., Zhu, C.B., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.M. and Floss, H.G. (1990). Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. *J. Bact.* **172**, 4816-4826.

Bassett, E.W. and Tanenbaum, S.W. (1960). Acetyl-coenzyme A in patulin biosynthesis. *Biochim. Biophys. Acta.* **40**, 535-537.

Beck, J., Ripa, S., Siegner, A., Schiltz, E. and Schweizer, E. (1990). The multifunctional 6-methylsalicyclic acid synthase gene of *Penicillium patulum* - its gene structure relative to that of other polyketide synthases. *European journal of Biochemistry* **182,** 487-498.

Belandreau, J. and Knowles. (1978). Interactions between Non-Pathogenic Soil Microorganisms and Plants, Elsevier, Amsterdam.

Bentley, R. and Bennett, J.W. (1999). Constructing Polyketides: From Collie to combinatorial biosynthesis. *Annu. Rev. Microbiol* **53,** 411-446.

Bevitt, D.J., Cortes, J., Haydock, S.F. and Leadley, P.F. (1992). 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.* **204,** 39-49.

Bibb, M.J., Biró, H., Motamedi, H., Collins, J.F. and Hutchinson. (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J.* **8,** 2727-2736.

Bibb, M.J., Sherman, D.H., Omura, S. Hopwwod, D.A. (1994). Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. *Gene* **142,** 31-39.

Birch, A.J. and Donovan, F.W. (1953). Studies in relation to biosynthesis. I. Some possible routes to derivatives of orcinol and phloroglucinol. *Aust. J. Chem.* **6,** 360-368.

Bisang, C., Long, P.F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.-L., Cox, R.J., Simpson, T.J., Staunton, J. and Leadley, P.F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401,** 502-505.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Ftzgerald, L.M., Clayton, R.A., Gocayne, J. D., Kerlavage, A.R., Dougherty, B.A., Tomb, J. F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Goeghagen, N.S.M. and Vebter, J.C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273,** 1058-1073.

Butler, M.J., Friend, E.J., Hunter, I.S., Kaczmarek, F.S., Sugden, D.A. and Warren, M. (1989). Molecular cloning of resistance genes and architecture of a linked gene cluster involved in biosynthesis of oxytetracycline by Streptomyces rimosus. *Mol. Gen. Genet.* **215,** 231-238.

Carreras, C.W., Pieper, R. and Khosla, C. (1997). The chemistry and biology of fatty acid, polyketide, and nonribosomal peptide biosynthesis. *Topics in Current Chemistry* **188,** 85-126.

Chater, K.F. and Bruton, C.J. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO J.* **4,** 1893-1897.

Colie, J.N. (1907). Derivatives of the multiple ketone group. *Proc. Chem. Soc.* **23,** 230-31.

Colie, J.N. (1907). Derivatives of the multiple keten group. *J. Chem. Soc.* **91,** 1806-1813.

Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadley, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-polyketide synthase of *Saccharopolyspora erythreae*. *Nature* **346,** 176-178.

Cresswell, N., Herron, P.R., Saunders,V.A. and Wellington, E.M.H. (1993). The fate of introduced streptomyces, plasmid and phage populations in a dynamic soil system. *JGM.* **138**, 656-666.

Cross, T. and Johnston, D.W. (1972). *Thermoactinomyces vulgaris*, distribution in natural habitats. In *Spore Research*, pp. 315-330. Edited by A.M. Barker, G.W. Gould and J. Wolf. Academic Press, London.

Davies, J. (1997). Drugs form dirt: New approaches to molecular diversity. Rhone – Poulenc – How best to generate leads from natural substances, Paris, October 1997.

Dietz, A. (1988). Practical and proposed cooperative investigational criteria for taxonomic studies of the Actinomycetales. In *Biology of Actinomycetes*, pp. 203-209. Editors Y. Okami, T. Beppu and H. Ogawara. Japan Scientific Societies Press, Tokyo.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J and Katz, L. (1991). Modular organisation of genes required for complex polyketide biosynthesis. *Science* **204,** 39-46.

Donadio, S. and Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111,** 51-60.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L. (1992). Biosynthesis of the erythromycin macrolactone and a rational approach for producing hybrid macrolides. *Gene* **115**, 97-103.

Edwards, U., Rogall, T., Blocker, H., Emde, M. and Bottger, C. (1989). Isolation and direct nucleotide determination of entire genes. Characterisation of a gene coding for 16S ribosomal RNA. *Nuc. Acids. Res.* **17**, 7843-7853.

Egan, S., Weiner, P., Kallifidas, D. and Wellington, E.M.H. (1998). Transfer of streptomycin biosynthesis gene clusters within streptomycetes isolated from soil. *Appl. Environ. Microbiol.* **64,** 5061-5063.

Eison, J. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: Comparison of trees of RecA and 16S rRNAs from the same species. *J. Mol. Evol.* **41**, 1105-1123.

Embley, T.M. and Stackebrandt, E. (1994). The molecular phylogeny and systematics of the actinomycetes. Annu. Rev. Microbiol. **48,** 257-289.

Ensign, J.C. (1992). Introduction to the actinomycetes. *In The Prokaryotes*, 2nd ed. Volume 1, pp. 811-815. Edited by A. Balows, H.G. Troper, M. Dworkin, W. Harder and K.-H. Schleifer. New York: Springer Verlag.

Epp, J.K., Huber, M.L., Turner, J.R., Goodson, T. and Schoner, B.E. (1989 Production of a hybrid macrolide antibiotic in *Streptomyces ambofaciens* and *Streptomyces lividans* by introduction of a cloned carbomycin biosynthetic gene from *Streptomyces thermotolerans*. *Gene* **85**, 283-301.

Feitelson, J.S. and Hopwood, D.A. (1983). Cloning of a *Streptomyces* gene for *O*-methyl-transferase involved in antibiotic biosynthesis. *Mol. Gen. Genet.* **190**, 394-398.

Felsenstein, J. (1993). PHYLIP (Phylogeny Interface Package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Felske, A., Rheims, H., Wolterink, A., Stackebrandt, E. and Akkermans, A.D.L. (1997). Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacter in grassland soils. *Microbiol.* **143**, 2983-2989.

Fernández-Moreno, M.A., Martín-Triana, A.J., Martínez, E., Niemi, J., Kieser, H.M., Hopwood, D.A. and Malpartida, F. (1992). *abaA*, a new pleiotropic regulatory locus for antibiotic production in *Streptomyces coelicolor*. *J. Bact.* **9**, 2958-2967.

Feitelson, J.S. and Hopwood, D.A. (1983). Cloning of a Streptomyces gene for O-methyltransferase involved in antibiotic biosynthesis. *Mol. Gen. Genet.* **190**, 394-398.

Fu, H., McDaniel, R., Hopwood, D.A. and Khosla, C. (1994). Engineered biosynthesis of novel polyketides: stereochemical course of two reactions catalyzed by a polyketide synthase. *Biochemistry*.**39**, 321-326.

Fuhrman, J.A., Comeau, D.E., Hagstrom, A. and Chan, A.M. (1988). Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Appl. Environ. Microbiol.* **54**, 479-483.

Goodfellow, M. and Williams, S.T. (1982). Ecology of actinomycetes. *J. Gen. Microbiol.* **136**, 19-36.

Goodfellow, M. and Dickinson, C.H. (1985). Delineation and description of microbial populations using numerical methods. In *Computer-Assisted Bacterial Systematics*, pp. 165-225. Edited by M. Goodfellow, D. Jones and F.G. Priest. London: Academic Press.

Gottlieb, D. (1976). The production and role of antibiotics in soil. *Journal of Antibiotics* **24**, 987-1000.

Gray, T.R. and Williams, S.T. (1971) *Soil Microorganisms.* Edited by D. Grothues and B. Tommler. Longman, London.

Grimm, A., Madduri, A., Ali, A. and Hutchinson, C.R. (1994). Characterisation of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151,** 1-10.

Hain, T., Ward-Rainey, N., Kroppenstedt, R.M., Stackebrandt, E. and Rainey, F.A. (1997). Discrimination of *Streptomyces albidoflavus* strains based on the size and number of 16S-23S ribosomal DNA intergenic spacers. *IJSB*, **47**, 202-206.

Hallam, S.E., Malpartida, F. and Hopwood, D.A. (1988). Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor. Gene* **74,** 305-320.

Hart, M.C., Elliot, G.N., Osborn, A.M., Ritchie, D. and Strike, P. (1998). Diversity amongst Bacillus *merA* genes amplified from mercury resistant isolates and directly from mercury polluted soil. *FEMS Microbiol. Ecol.* **27**, 73-84.

Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortes, J. and Leadley, P.F. (1991). Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea* sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* **230,** 120-

Head, I.M., Saunders, J.R. and Pickup, R.W. (1998). Microbial evolution , diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb. Ecol.* **35**, 1-21.

Herlt, A.J., Pickards, R.W. and Wu, J.P. (1985). The structure of streptonigrin and a comment on the biosynthesis of the streptonigrin antibiotics. *J. Antibiot.* **38,** 516-518.

Herrick, B.J., Madsen, E.L., Batt, C.A. and Ghiorse, W.C. (1993). Polymerase chain reaction amplification of naphthalene catabolic and 16S rRNA gene sequences from indigenous sediment bacteria. *Appl. Environ. Microbiol.* **56,** 1406-1412.

Herron, P.R. and Wellington, E.M.H. (1990). New method for extraction of streptomycete spores from soil and application to the study of lysogeny in sterile amended and nonsterile soil. *Appl. Environ. Microbiol.* **56**, 1406-1412.

Heuer, H., Krsek, M., Baker, P., Smalla, K. and Wellington, E.M.H. (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16SrRNA and gel-electrophoretic separation in denaturing gradients. *Appl. Environ. Microbiol.* **63**, 3233-3241.

Hiltner, L. (1904). *Arb. Dtsch. Landwirt. Ges.* **98**, 59.

Holben, W.E., Jansson, J.K., Chelm, B.K. and Tiedje, J.M. (1988). DNA probing method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbio.* **54**, 703-711.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985). Genetic manipulations of *Streptomyces*. A laboratory manual. The John Innes Foundation, Norwich.

Hopwood, D.A., Malpartida, F.; Kieser, H.M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B.A.; Floss, H.G. and Omura, S. (1985). Production of hybrid antibiotics by genetic engineering. *Nature* **314**, 642-644.

Hopwood, D.A. and Sherman, D.H. (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24,** 37-66.

Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. *Chemical Reviews* **97**, 2465-2497.

Huddleston, A.S., Cresswell, N., Cristina, M., Neves, P., Beringer, J.E., Baumberg, S., Thomas, D.I. and Wellington, E.M.H. (1997). Molecular detection of streptomycin-producing streptomycetes in Brazilian soils. *Appl. Environ. Microbiol.* **63,** 1288-1297.

Hugenholtz, P. and Pace, NR. (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *TIBTECH* **14**, 190-197.

Hutchinson, C.R., Decker, H., Motamedi, H., Shen, B., Summers, R.G., Wendt-Pienkowski, E. and Wessel, W.L. (1993). Molecular genetics and biochemistry of a bacterial type II polyketide synthase. In *Industrial Microorganisms: Basic and Applied Molecular Genetics.* Chapter 26. Edited by R.H. Baltz, G.D. Hegeman and P.L. Skatrud. American Society for Microbiology.

Hutchinson, C.R. and Fujii, I. (1995). Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **49,** 201-238.

Hutcinson, C.R. (1999). Microbial polyketide synthases: More and more prolific. *Proc. Natl. Acad. Sci. USA* **96,** 3336-3338.

Itoh, Z., Suzuki, T., Nataya, M., Inoue, M. and Mutsuhashi, S. (1984). Gastrointestinal motor-stimulating activity of macrolide antibiotics and analysis of their side effects on the canine gut. *Antimicrob. Agents Chemother.* **26,** 863-869.

Jacobsen, C.S. and Rasmussen, O.F. (1992). Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl. Environ. Microbiol.* **58**, 2458-2462.

Jones, A.L. and Keller, U. (1997). Biochemistry and gentetics of actinomycin production. In *Biotechnology of antibiotics*. Edited by W.R. Strohl. Marcel Dekker, Inc.

Katz, L. and Donadio, S. (1993). Polyketide synthesis: Prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* **47**, 875-912.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. and Cane, D.E. (1999). Tolerance and specificity of polyketide synthases. **Annu. Rev. Biochem. 68,** 219-253.

Kim, C.-G., Kirsching, A., Bergon, P., Zhou, P., Su, E., Sauerbrei, B., Ning, S., Ahn, Y., Breuer, M., Leistner, E. and Floss, G.G. (1996). Biosynthesis of 3-amino-5-hydroxybenzoic acid, the precursor mC7N units in ansamycin antibiotics. *J. Am. Chem. Soc.* **118,** 7486-7491.

Kirshtein, J.D., Paerl, H.W. and Zehr, J.P. (1991). Amplification, cloning, and sequencing of a *nifH* segment from aquatic microorganisms and natural communities. *Appl. Environ. Microbiol. 57,* 2645-2650.

Klijn, N., Weerkamp, A.H. and de Vos, W.M. (1991). Identification of mesophilic lactic acid bacteria by using polymerase chain reaction-amplified variable regions of 16S rRNA and specific DNA probes. *Appl. Environ. Microbiol.* **57**, 3390-3393.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.S., Rehrauer, W.M. (1994). Biochemistry of homologous recombination in *Escherichia coil. Microbiol Rev.* **58**, 401-465.

Lan, R. and Reeves, P. (1992). Gene transfer is a major factor in bacterial evolution. *FEMS Microbiol. Lett.* **100**, 509-516.

Lechavalier, H.A. (1964). Principles and application in aquatic microbiology. In *The Actinomycetes*, pp. 230-250. Edited by Heukelekan and Dondero. John Wiley and Sons, New York.

Lloyd, A. and Sharp, P.M. (1993). Evolution of the *recA* gene and the molecular phylogeny of bacteria. *J. Mol. Evol.* **37**, 399-407.

Ludwig, W., Neumaier, J., Klugbauer, N., Brockman, E., Roller, C., Jilg, S., Reetz, K., Schachtner, I., Ludvigsen, A. and Bachleitner. (1993). Phylogenetic relationships of bacteria based on comparative sequence analysis of elongationfactor Tu and ATP-synthase beta-subunit genes. *Ant. Van. Leeuwen.* **64**, 285-305.

Ludwig, W. and Schleifer, K.H. (1994). Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiology Reviews* **15**, 155-173.

Lynch, J.M. and Painting, L.M. (1980). Cultivation and the soil biomass. *Soil. Biol. Biochem.* **12**, 29-33.

Magnuson, K., Jackowski, S., Rock, C.O. and Cronan, J.E. (1993) Regulation of fatty acid biosynthesis in Escherichia coli. *Microbiol. Rev.* **57,** 522-542.

Maidak, B.L., Larson, N., McCaughey, M.J., Overbeek, R., Olsen, G.J., Fogel, K., Blandy, J. and Woese, C.R. (1994). The Ribosomal Database Project. *Nucleic Acids Res.* **22**, 3485-3487.

Malpartida, F. and Hopwood, D.A. (1984). Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462-464.

Malpartida, F. and Hopwood, D.A. (1986). Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2).*Mol. Gen. Genet.* **205,** 66-73.

Malpartida, F., Hallum, S.E., Kieser, H.M., Motamedi, H., Hutchinson, C.R., Butler, M.J., Sugden, D.A., Warren, M., McKillop, C., Bailey, C.R., Humphreys, G.O. and Hopwood, D.A. (1987). Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes. *Nature* **325,** 818-821.

Malpartide, F., Hallam, S.E., Keiser, H.M., Motamedi, H., Hutchinson, C.R., Butler, M.J., Sugden, D.A., Warren, M., McKillop. C. and Bailey, C.R. (1987). Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthesis genes. *Nature* **325**, 818-821

Marsh, P., Morris, N.Z. and Wellington, E.M.H. (1998). Quantitative molecular detection of *Salmonella typhimurium* in soil and demonstration of an active but non-culturable population. *FEMS Microbiology Ecology* **27**, 351-363.

Martin, J.F. and Liras, P. (1976). Rapid incorporation of precursors into candicidin by resting cells of *Streptomyces griseus. J. Antibiot* **29,** 1306-1309.

McCarthy, A.D. and Hardie, D.G. (1984). Fatty-acid synthases - an example of protein evolution by gene fusion. *Trends Biochem. Sci.* **9,** 60-67.

McDAniel , R.S., Ebert-Khosla, S., Hopwood, D.A. and Khosla, C. (1994). Engineered biosynthesis of novel polyketides: actVII and actIV encode aromatase and cyclase enzymes respectively. *J. Am. Chem. Soc.* **116**, 10855-10859.

McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. and Khosla, C. (1995). Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375,** 549-554.

McDonald, I.R. and Murrell, J.C. (1997). The methanol dehydrogenase structural gene *mxaF* and its use as a functional gene probe for methanotrophs and methylotrophs. *Appl. Environ. Microbiol.* **63,** 3218-3224.

Metsä-Ketelä, M., Salo, V., Halo, L., Hautala, A., Hakala, J., Mäntsälä, P. and Ylihonko, K. (1999). An efficient approach for screening minimal PKS genes from Streptomyces. *FEMS Microbiology Letters* **180**, 1-6.

Miller, R.V. (1992). *recA. In Encyclopedia of Microbiology*, pp. 509-517. Edited by J. Lederberg, M. Alexander, D.A. Hopwood, B.H. Igleweskl, A.I. Laskin. Academic Press, Inc.

Mirza, M.S., Hameed, S. and Akkermans, A.D.L. (1994). Genetic diversity of *Datisca cannabina*-compatible Frankia strains as determined by sequence analysis of the PCR-amplified 16S rRNA gene. *Appl. Environ. Microbiol* **60,** 2371-2376.

Monaghan, R.L. and Tkacz, J.S. (1990). Bioactive microbial products: focus on mechanism of action. *Annu. Rev. Microbiol.* **44,** 271-301.

Morris, N.Z., Standage, S. and Wellington, E.M.H. (1998). DNA extraction from spores in soil. In *Molecular Ecology Manual.* 4th edn, 1.3.7. Edited by A.D.L. Akkermans, J.D. van Elsas and F.J. de Bruijn. Kluwer Academic Publishers.

Motamedi, H. and Hutchinson, C.R. (1987). Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumour antibiotic of *Streptomyces glaucescens. Proc. Natl. Acad. Sci. USA* **84**, 4445-4449.

Motamedi, H., Cai, S.J., Shafiee, A. and Elliston, K.O. (1997) Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506. *Eur. J. Biochem.* **244**, 74-80.

Mullis, K.B. and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155,** 355-351.

Muyzer, G., de Waal, E.C. and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes for 16S rRNA. *Appl. Environ. Microbiol.* **59,** 695-700.

Nuβbaumer, B. and Wohlleben, W. (1994). Identification, isolation and sequencing of the *recA* gene of *Streptomyces lividans* TK24. *FEMS Microbiol. Lett.* **118**, 57-64.

Ogram. A., Sayler, G.S. and Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* **7**, 57-66.

Ohkuma, M. and Kudo. T. (1996). Phylogenetic diversity of the intestinal bacterial community in the Termite *Reticulitermes speratus*. *Appl. Environ. Microbiol.* **62**, 461-468.

Okami, Y. and Hotta, K. (1988). Search and discovery of new antibiotics. In *Actinomycetes in Biotechnology*, pp. 33-67. Edited by M. Goodfellow, S.T. Williams and M. Mordarski. Academic Press. London.

Olsen, G.J., Lane, D.J., Giovannoni, S.J. Pace, N.R. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Ann. Rev. Microbiol.* **40**, 337-365.

Olsen, G.J. and Woese, C. (1993). Ribosomal RNA: a key to phylogeny. *The FASEB journal* **7**, 113-123.

Omura, S. (1992). The expanded horizon for microbial metabolites a review. *Gene* **115**, 141-149.

Øvreås, L., Forney, L., Daae, F,L. and Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **63**, 3367-3373.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**, 1568-1571.

Paul, E.A. and Clark, F. E. (1989). *Soil Microbiology and biochemistry*. Academic Press Inc, Ca. USA.

Porteous, L.A. and Armstrong, J.L. (1988). Recovery of bulk DNA from soil by a rapid small-scale extraction method. *Curr. Microbiol.* **22**, 345-348.

Queener, S.W. and Lively, D.H. (1986). Screening and selection for strain improvement. In *Manual of Industrial Microbiology and Biotechnology*, pp. 155-169. Edited by A.L. Demain and N.A. Solomon. American Society for Microbiology.

Rheims, H., SprÖer, C., Rainey, F.R. and Stackebrandt. (1996). Molecular biological evidence for the occurrence of uncultured members of the actinomycete line of descent in different environments and geographical locations. *Microbiology* **142**, 2863-2870.

Rothrock, R.J., Wahab, F.A. and Sundrum, J. (1984). Role of antibiotics in the antagonism of *Streptomyces hygroscopicus* var. *geldanus* to *Rhizoctonia solani* in soil. *Can. J. Microbiol* **30**, 1440-1446.

Ruan , X., Stassi, D., Lax, S.A. and Katz, L (1997). A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 29253, a rapamycin-producing strain. *Gene* **203**, 1-9.

Salama, M., Sandine, W. and Giovanni, S. (1991). Development and application of oligonucleotide probes for identification of Lactococcus. *Appl. Environ. Microbiol.*. **57**, 1313-1318.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloninf: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.

Schupp, T., Toupet, C., Engel, N. and Goff, S. (1998). Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from Amycolatopsis mediterranei. *FEMS Microbiol. Lett.* **159**, 201-207

Schwecke, T., Aparicio, J.F., Moln·r, I., Konig, A., Khaw, Haydock, S.F., Oliynyk, M., Caffrey, P., Cortes, J.B., Bohm, G.A., Staunton, J. and Leadley, P.F. (1995). The biosynthesis gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci.* USA **92,** 7839-7843.

Schweizer, M., Roberts, L.M., Holtke, H.-J., Takabayashi, K., Hollerer, E., Hoffmann, B., Muller, G., Kottig, H. and Schweizer, E. (1987). The pentafunctional FASI gene of yeast: its nucleotide sequence and order of catalytic domains. *Mol. Gen. Genet.* **203,** 479-486.

Seow, K-T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. and Davies, J. (1997). A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bact.* **179**, 7360-7368.

Shen, B., Summers, R.G., Wendt-Pienkowski, E. and Hutchinson, C.R (1995). The *Streptomyces glaucescens tcmKL* polyketide synthase gene and *tcmN* polyketide cyclase genes givern the size and shape of aromatic polyketides. *J. Am. Chem. Soc.* **117**, 6811-6821.

Sherman, D.H., Malpartida, F., Bibb, M.J., Keiser, H.M., Bibb, M.J. and Hopwood, D.A. (1989). Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of Streptomyces violaceoruber TY22. *EMBO.* **8**, 2717-2725.

Siggaard-Anderson, M. (1993). *Protein Sequencing Data Anal.* **5**, 325.

Sneath, P.H.A. (1989). Numerical taxonomy. *In Bergey's Manual of Systematic Bacteriology* Volume 4, pp. 2303-2305. Edited by S.T. Wiliams, M.E. Sharpe and J.G Holt. Baltimore: Williams and Wilkins.

Somerville, C.C., Knight, I.T., Straube, W.L. and Colwell, R.R. (1989). Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl. Environ. Microbiol.* **55**, 548-554.

Soulides, D.A. (1965). Antibiotics in soils: VII. Production of streptomycin and tetracyclines in soil. *Soil Science*. **100**, 200-205.

Stackebrandt, E., Ludwig, W., Seewaldt, E. and Schleifer, K. (1983). Phylogeny of sporeforming members of the order *Actinomycetales*. *International Journal of Systematic Bacteriology* **33,** 173-180.

Stackebrandt, E., Witt, W., Kemmerling, C., Kroppenstadt, R. and Liesack. (1991). Designation of streptomycete 16S and 23S rRNA-based target regions for oligonucleotide probes. *Appl. Environ. Microbiol.* **57, 1468-1477.**

Stackebrandt, E. (1992). Molecular microbial ecology. In *Encyclopedia of Microbiology*, pp 171-179. Edited by J. Lederberg, M. Alexander, D.A. Hopwood, B.H Igleweski and A.I. Laskin. Academic Press, Inc.

Stackebrandt, E., Liesack, W. and Witt, D. (1992). Ribosomal RNA and rDNA sequence analysis. *Gene*. **115**, 255-260.

Stackebrandt, E. and Liesack, W. (1993). Nucleic acids and classification. In *Handbook of New Bacterial Systematics*, pp. 151-194. Edited by M. Goodfellow and A.G. O'Donnell. London: Academic Press.

Stackebrandt, E., Rainey, F.A. and Ward-Rainey, N.L. (1997). Proposal for a new hierarchic classification system, Actinobacteria classis nov. *International Journal of Systematic Bacteriology* **47**, 479-491.

Standage, S. (1998). Molecular, genetic and ecological analysis of gentamicin-producing *Micromonospora* species. Ph.D thesis, University of Warwick.

Staunton, J. and Wilkinson, B. (1997). Biosynthesis of erythromycin and rapamycin. *Chem. Rev.* **97**, 2611-2629

Steffan, R.J., Goksoyr, J., Bej, A.K. and Atlas, R.M. (1988). Recovery of DNA from soils and sediments. *Appl. Environ. Microbiol.* **54,** 2908-2915.

Steffan, R.J. and Atlas, R.M. (1988).DNA amplification to enhance detection of genetically engineered bacteria in environmental samples. *Appl. Environ. Microbiol.* **54**, 2185-2191.

Steffan, R.J. and Atlas, R.M. (1991). Polymerase chain reaction.: Applications in environmental microbiology. *Annu. Rev. Microbiol.* **45**, 137-161.

Stone, M.J. and Williams, D.H. (1992). On the evolution of functional secondary metabolites (natural products). *Molecular Microbiology* **6,** 29-34.

Stotzky, G. (1997). Soil as an environment for microbial life. In *Modern Soil Microbiology*. Edited by J. D. van Elsas, J.T. Trevors and E. M. H. Wellington. Marcel Dekker, Inc.

Strohl, W.R. (1997). Industrial antibiotics: Today and the future. In *Biotechnology of antibiotics*. Edited by W.R. Strohl. Marcel Dekker, Inc.

Straub, T.M., Pepper, I.L., Abbaszadegan, M. and Gerba, C.P. (1994) A method to detect enteroviruses in sewage sludge-amended soil using PCR. *Appl. Environ. Microbiol.* **60**, 1014-1017.

Summers, R.G., Ali, A., Shen, B., Wessel, W.A. and Hutchinson, C.R. (1995). Malonyl-coenzyme A:Acyl carrier protein acyltransferase of *Streptomyces glaucescens*: a possible link between fatty acid and polyketide biosynthesis. *Biochemistry* **34,** 9389-9402.

Tanner, M.A., Goebel, B.M., Dojka, M.A. and Pace, N.R. (1998). Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* **64**, 3110-3113.

Tebbe, C.C. and Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and yeast. *Appl. Environ. Microbiol.* **59**, 2657-2665.

Teske, A., Sigalevisch, P., Cohen, Y. and Muyzer. (1996). Molecular identification of bacteria from coculture by denaturing gradient gel electrophoresis of 16S ribosomal DNA fragments as a tool for isolation in pure cultures. *Appl. Environ. Microbiol.* **62**, 4210-4215.

Thompson, C.J., Ward, J.M. and Hopwood, D.A. (1982). Biochemical characterization of resistance determinants cloned from antibiotic-producing streptomycetes.*J. Bacteriol.* **151,** 668-685

Torsvik, V., Goksoyr, J. and Dane, F.L. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**, 782-787.

Torsvik, V., Daae, F.L., Sandaa, R.A. and OvreÁs, L. (1998). Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* **64**, 53-62.

Tsai, Y.L and Olsen. B.H. (1991). Rapid method for direct extraction of DNA from soil and sediments. *Appl. Environ. Microbiol* **57**, 1070-1074.

Tsoi, C.J. and Khosla, C. (1995). Combinatorial biosynthesis of 'unnatural' natural products: the polyketide example. *Chem. Biol.* **2,** 355-362.

Tuan, J.S., Weber, J.M., Staver, M.J., Leung, J.O., Donadio, S. and Katz, L. (1990). Coning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid *Gene* **90,** 21-29.

Vanden Boom, T.J. and Cronan, J. E. (1989). Genetics and regulation of bacterial lipid metabolism. *Annu. Rev. Microbiol* **43,** 317-343.

Walia, S., Khan, A. and Rosenthal, N. (1990). Construction and applications of DNA probes for detection of polychlorinated biphenyl-degrading genotypes in toxic organic-contaminated soil environments. *Appl. Environ. Microbiol.* **56,** 254-259.

Wallace, K.K., Reynolds, K.A., Koch, K., McArthur, H.A.I., Brown, M.S., Wax, R.G. and Moore, B.S. (1994). Biosynthetic studies of ascomycin (FK520): formation of the (1R,3R,4R)-3,4-dihyddroxycyclohexanecarboxylic acid-derived moiety. *J. Am. Chem. Soc.* **116,**11600-11601.

Wang, Y., Zhang, Z. and Ramanan, N. (1997). The actinomycete *Thermobispora* contains two distinct types of transcriptionally active 16S rRNA genes. *J. Bact.* **179**, 3270-3276.

Ward, N., Rainey, F.A., Goebel, B. and Stackenbrandt, E. (1995). Identifying and culturing the "unculturables": A challenge for microbiologists. *In Microbial and Ecosystem Function*, 89-110. Edited by D. Allsopp, R.R. Colwell and D.C. Hawksworth. CAB International.

Weiner, P., Egan, S. and Wellington, E.M.H. (1998). Evidence for transfer of antibiotic-resistance genes in soil populations of streptomycetes. *Molecular Ecology* **7,** 1205-1216.

Wellington, E.M.H. and Williams, S.T. (1978). Preservation of streptomycete inoculum in frozen glycerol. *Microbios. Letters.* **6**, 151-157.

Wellington, E.M.H. and Cross, T. (1983). Taxonomy of antibiotic-producing actinomycetes and new approaches to their selective isolation. *Prog. Industr. Microbiol.* **17,** 7-36.

Wellington, E.M.H. and Toth, I.K. (1983). Actinomycetes. In *Methods of soil analysis, part 2 Chemical and microbiological properties* -Agronomy Monograph no. 9 (3rd edition ASA-SSSA book series no. 5 pp. 269-290).

Wellington, E.M.H., Stackebrandt, E., Sanders, D., Wolstrup, J. and Jorgensen, N.O.G. (1992). Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces Waksman* and *Henrici* 1943, 339^{AL}. *International Journal of Systematic Bacteriology* **37,** 463-464.

Wellington, E.M.H. and Toth, I. (1994). Actinomycetes. In: Methods of soil analysis. Soil Society of America, Madison, Wisconsin.

Weyland, H. (1981). Distribution of actinomycetes on the sea floor. *Zentralbl. Bakteriol. Suppl.* **11**, 185-193.

Williams, S.T. (1982). Are antibiotics produced in soil? *Pedobiologica.* **23**, 427-435.

Willams, S. T. and E. M. H. Wellington. (1982). Actinomycetes. pp969-987. In: Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties. 2 nd Edition (Eds. A. L. Page, R. H. Miller, and D. R. Keeney). American Soceity of Agronomy and Soil Science Soceity of America, Madison, Wisconsin.

Williams, S.T., Goodfellow, M., Wellington, E.M.H., Vickers, J,C., Alderson, G., Sneath, P.H.A., Sackin, M.J. and Mortimer, A.M. (1983). A probability matrix for the identification of some streptomyces. *Journal of General Microbiology* **129,** 1815-1830.

Wintzingerode, F.V., GÖbel, U.B. and Stackebranbt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**, 213-229.

Witkowski, A., Rangan, V.S., Randhawa, Z.I., Amy, C.M. and Smith, S. (1991). Structural organization of the multifunctional animal fatty acid synthase. *Eur. J. Biochem.* **198,** 571-579.

Witt, D. and Stackebrandt, E. (1990). Unification of the genera Streptoverticilleum and *Streptomyces Waksman* and *Henrici* 1943, 339^{AL}. *Systematic and Applied Microbiology* **13,** 361-371.

Woese, C. (1987). Bacterial evolution. *Microbiol . Rev.* **51**, 2210271.

Woodward, R.B. (1957). Struktur und Biogenese der Makrolide. *Angew. Chem.* **69**, 50-58.

Yap, W.H., Zhang, Z. and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire operon. *J. Bact.* **181**, 5201-5209.

Ye, J., Dickens, M.L., Plater, R., Li, Y., Lawrence, J. and Strohl, W.R. (1994)..Isolation and sequence analysis of polyketide genes from the daunomycin-producing *Streptomcyes* sp. Strain C5. *J. Bacteriol.* **176,** 6720-6280

Yu, T,-W., Shen, Y., McDaniel, R., Floss, H.G., Khosla, C., Hopwood, D.A. and Moore, B. (1998). Engineered biosynthesis of novel polyketides from Streptomyces spore pigment polyketide synthases. *J. Am. Chem. Soc.* **120,** 7749-7759.

Chapter 9

Appendices

9.1. Sequences generated from this study

9.1.1. Cuban actinomycetes, 16S rDNA sequences :

>NV1 (*Streptomyces chartreusis*, 95 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATAACACTCCTGTCCTCCTGGACGGGGGTTAAAATCTCCGGCGGTGAAGGATTAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAATGGCTCACCAAGGCTACAACGGGTACCCGGCCTGAAAGGGCGACCGGCCACACTGGGA
CTGAAACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCCTAAAGCCTGATGCACT
ACCCGCGTGAGGGATGACCGCCTTCCGGTTGTAAACCACTTTCACAAGGA
>NV2 (*S. avermitilis*, 96 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATAATACCCTCGCAGGCATCTGCGAGGGTTAAAAGCTCCGGCGGTGCAGGATGAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAATGGCTCACCAAGGCTACTACGGGTAGCCGGCCTGAAAGGGCAACCGGCCACACTGGGA
CTGAAACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCGGC
GACCCGCGTAAGGGATAACGGCCTTCGGGTTTTAAACCTCTTTCACAGGGA
>NV3 (*Streptomyces chartreusis*, 98%)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATAACACTCCTGTCCTCCTGGACGGGGGTTAAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACAACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGG
ACTGAAACACGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCCAATGGGCGAAAGCCTGATCCAGCG
ACCCCCCTGAGGGATAACGGCTCCGGGTTGTGACCTCTTGCGCGGGGAAAGA
>NV6 (*Streptomyces* spp., 95 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCTGGGAGGCATCTCCCGGGTTGTAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATCA
TCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACAACGGGTACCGGCCTGAGAGGGCGACCGGCCACGCTGGGA
CTGAAACACGGCCCGTACTCCTACGGGAGGCAGCACTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCACC
GACCCCGCGTGAGGGATAACAGCCTTCTGGTTGTTAACCTCTTGCCCGGGGA
>NV7 (*Streptomyces* spp., 93 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCTGGGAGGCATCTCCCGGGTTGTAAAGCTCCGGCGGTGAAGGATTGTCCCACGGCCATTCC
CCTTGTTGGTGAGGTATCGGCTCACCTAGGAGACAACGGGTGGCTGCCTGATACGGCCACCGGCCCCTCTCGGAC
TTAAACCCGCACCAACGCCTCCTGGAGGCCCCTGTGGGGATTATTCCCCGATGGGCCAAACCCGATCCCTCCACTC
CCCTCACGGACTCCACCCTCCGGTCTTTCCACCCTTTCCAGCTGGGAAA
>NV8 (*S. pseudovenezuelae*, 95 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACCACTCCCCCTCGCATCGGTGGGGGTTGAAAGCTCCGGCGGTGAAGGATGATCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACAACGGGTATCCGGCCTGAAAGGGCGACCGGCCACACTGGG
ACTGAGACACGGCCCAAACTCCTACGGGAGGCATCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAT
CTACCCGCGTGAGGGATGACGGCCTCCGGGTTCTAAACCTCTTGCCCCGAGG
>NV9 (*S. pseudovenezuelae*, 94 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACCACTCCCACTCGCATGGGTGGGGGTTGAAAGCTCCGGCGGTGAAGGATGAACCCGCGGCCTATCA
CCTTGTTGGTGAGGTACTGGCTCACCAAGGCTACAACGGGTATCCGGCCTGAAAGGGCAACCGGCCGCACTGGGA
CTGAAACACGGCCCATACTCCTACGGGAGGCACCACTGGGGAGTATTGCACAGTGGGCCAGAACCTTGATCCCCC
GACCCCCCTTGAGGGATTACCGCCTTCCGGGTTTTAAACCTCTCTCACCC
>NV10 (*S. purpurascens*, 90 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCGGCACTCTGGGACAATCCCTGGAAACGGGGTCTAA
TACCGGATATTGACCATCTTGGGCATCCTTGATGGTGGAAAGCTCCGGCGGTGCAGGATGAGCCCTCGGCCTATCT
CCTATTTGGTGACGTAATGGCTCACCAAGGCCATGAGGGGTACCCGGCCTGAAAGGGCTACCGGCCACGTCTGGG
ACTGAAACACGGTCCAAACGCCCACGGGAAGCACCATTGGGGAGTATTGCTCAATGGGCCAAACCCTTTATGCGC
CCACGCCCCGTTAGGGATGAGCGCCCCCCGGTTTTAAACCCCTTTCGCC
>NV11 (*Streptomyces spp.*, 92 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATAACACTTCGCACCGCATGGGGGAGGGTGGAAAGCTCCGGCGGTGCAGGATGACCCCGCGGTCTAGCT
GCTTGTTGGTGGGGTGATGGCCTACCATGGCGACGACGGGTAGCCGGCCTGAAAGGGCGACCGGCCGGTCTGGGA
CTGAAACCGGCCATACTCCTACCGGAGGCATTATTGGGGAATTTTCCCCCCCCCGGTCACAACCTGTCCCCCCACCC
CCTCAGGGAAGTCTCCCCCCGGTTTTGCCCCCTCTACCACGCAACAACC
>NV12 (*S. purpurascens*, 96 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAATGGGGTCTAA
TACCGGATATTGACCATCTTGGGCATCCTTGATGGTGGAAGGCTCCGGCGGTGCAGGATGATCCCTCGGCCTGTCA
ACTGTTTGGTCACGTATCGGCTCCCCAAGGCCACAAGGGGTATGCGGCCTTACAGGGGTAGCGGCCTCGCTGGGA
CTGAAACTCGCTCATACGCCTGCGGGAGGCACTACTGTGGATTATTGAACCATTGATTGAAAGCCCTGAGTCCCCC
ACCCATTTAAGGATTACTGCCCCCGGGTGTTAACCCCTCTCTGCCGGG
>NV13 (*S. panayensis*, 93 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTCCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATCACTTCCTGAGGCATCTCGGGGGGTTGAAAGCTCCGGCGGTGGAGGATGAGCCCGCGGCCTATCA
CCTTGTTGGTGAGGTAACGGCTCACCAAGGCTACAACGGGTTCCCGGCCTGAAAGGGCGACCGGCCACACCTGGGA

CTGAAACACGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCACGA
CCCGCTTGAGGGATTACGCCTTCGGGTTGTAAACCTCTTTCACAGGGAAAA
>NV18 (*Streptomyces* spp., 97 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCTGGGAGGCATCTCCCGGGTTGTAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACAACGGGTAGCCGGCCTGAAAGGGCGACCGGCCACGCTGGG
ACTGAGACACGGCCCATACTCCTACGGGAGGCATCAGTGGGGAATATTGCTCAATGGGCGAAAGCCTGATGCGCC
AACCCCTTGAGGGATGACCGCCTCCTGGTTGTGAACCCTTTCGCAGGGAAA
>NV19 (*S. pseudovenezuelae*, 91 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACCACTGCCTCTCGCATGGGTGGGGGTTGAAAGCTCCGGCGGTGAAGGATGAACCCGCGGCCTATCA
CCTTGTTGGTGATGTATTGGCTCACCAAGGCGACGACGGGTGACCGGCTGAAAAGGCGACCCGCCACACTGGGAC
TGAAACACCGGCCAAACTCCTACGGGAAGCGACATGGGGGAATATTACACATGGGCGAAAACCTTGATTCACCAC
CCCCCTTAAAGGATAACCGCTTCCCGGTTTTTGACCCCTTTCGTCAGGGA
>NV20 (*Streptomyces* spp., 94 %)
TTAGTGGCGAACGGGTGAGTAACTCGTGGGCAATCTGCCCTTCACTCTGGGACAACCCCTGGAAACGGGGTCTAA
TACCGGATATGACCCTGGGATGCATCTCCCGGGTTGTTAATCGCCGGCGGTGAATGATTACCCCCCCGCCTATCTT
CTTTTTGGATGAAGTTCCGCCCCCCTTGACAGAACCGGTTCCCGCCTGAAAAGGGCAACCGGCCCCTCTTGGAATC
AAAACACCCTCCCCCACCCTCCGGGAAGGCTACCATGGGGGAATTTCTCCCCTTGGCCAAAACCCGTTTTCTTCAA
CCCACCTTTGTGGATAAAAACGCCTCCCCGTTCTGTCCCCCTGGCCC
>NV22 (*Streptomyces* spp., 95 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACGATCCTCCCGGGCATCTGGGAGGTTGGAAAGCTCCGGCGGTGCATGATTAACCCGCGGGCTATCA
CTTGTTGGTGAAGTAACGGGTCCCAAGGCCACAGGGGTTTCCCGCTGACAAGGCTATCCGGCCCCTTGGACTGAA
AGAGCGGCCCAACCCTACGGGAAGCACCAGTGGGGAATTATTCCCAATTGGCGAAAAACCATAATCCCCCACCCC
CCCTTAAAGGAATAACGGCCTCCCGTTTGTTTGACCCCCTTCGCCAGGGA
>NV24 (*Streptomyces* spp., 93 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACCACCTCCCGAGGCATCTTGGGGGGTTGAAAGCTCCGGCGGTGAAGGATTAGCCCGCGGCCTATCA
TCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACTGGTAACCGGCCTGAGAAGGCGACCGGCCGCGCTGGGA
CTGAAACACGGCCAAACTCCTACGGGGAAGCACAGTGGGGAATATTGCCAATTGGCCAAATCCTTATCTCCCACC
CCCCCTAAAAGGATTAACGGCCTCCGGTTATTTAACACCTTTCCCAGGGA
>NV25 (*S. pseudovenezuelae*, 97 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACCACTGCCACTCGCATGGGTGGGGGTTGAAACCTCCTGCGGTGAAAGATGAGCCCCCACCATCCTC
CTGTTGGTAAAGTCTGGCTCCCCTCGTCAACACCGGTTCTCCGCCGTAAAAGCCTACCCCCCCCACCCCGAAATAAA
AAACACCCCCATCCCCCCCCCGGAGGAGCCCCTGGAGCAAAATTTTTTTATTTGGCCATATCTCCTTCTCTCTCCACC
CTCCTCCTTGAAATAAAACCCCCCCTTCTTTCCCCCCCCCCCCCCCGC
>NV34 (*Streptomyces* spp., 96 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCTGGGAGGCATCTCCCGGGTTGTTAAGCTCCTGCGGTGAAGGATTATCCCGCGGCCTATCA
ACTTGTTGGTGAGGTCACGGCCCACCCGGCGACCACCGGTTACCCGCCGGAAAAGGCCACCCGGCCTCGCTGGGA
CTGAAACACCGCCCCCACTCCCCCCGGACGCCTCTGTGGGGGATATTTTACTCATAGGCGAGAACCCCATTGTCTC
CCAACCCCCCCTTGATGGGAATTAAGAGCCCTCTCGGTTGTTCTGCCCC
>NV42 (*S. longisporus*, 96 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCCGGGAGGCATCTCCCGGGTTGTAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGTGCCGGCCTGAAAAGGCGACCGGCCACACTGGGA
CTGAAACACGGACCATACTCCTACGGGAAGCATCAGTGGGGGAATATTGCACAATGGGCGCAACCTGATGCACCA
CCCCCTTTAAAGGATAACCGGCTTCCGGTTTTTAACCCCCTTCCCACGGAA
>NV43 (*Streptomyces* spp., 88 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTCCACACTGGGACAATCCCTGGAAACGGGGTCTAA
TACCGGATACCAGCCTCCCGGGCATCTGGGAAGTTGGAAACCTCCAGCGGTACATGATTAACCCACGGGCAATCA
GCTTGTTGGTTAGGTAGCGGCTCACCAGGGCGACAACGGGTATACGGCCTGACAAGGCGACGGGTACTCTGGGAC
TGAGACACTGCCCCGACTCCTACAGGACGCGGGATGGGGAATTTTGCAGTCTCCGCCAAAGCCTGATCCTTTCACC
CCCCCTTAAGGATTATGTCTTTCCGGTTTTCTTACTCTTCCCCCCGGAA
>NV49 (*S. cinnabarinus*, 96 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATATCACTCTCGCAGGCATCTGTGAGGGTCGAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATCA
GCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGTTCCGGCCTGATATGGCAACCGGCCACACTGGGAC
TGAAACACAGCCCACACTCCTACGGGAGGCAGTGGGGAATATTTGCACAATGGGCGAAAGCCTGATTCGCCCA
CCCCCCTTGAGGGATCACCGCCTTCCGGTTCTTAACTCTTTCACAGGGAA
>NV53 (*S. bikiniensis*, 92 %)
TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACGACCTGGGAAGGCATCTTCTCGGGTGGAAAGCTCCGGCGGTGAAGGATTATCCCGCGGCCTATCA
ACTTGTTGGTGAGGTAACGGCTCACCAAGGCGACAACAGGTATCCAGCCTGACAAGGGCACCCGCCACACTGGGG
ACTGAAACACCGCCCCCACTCCTACGGGAAGCTCCATGGGGAATATTTCCTATGGCGAAAACCTGATACCCCACC
CCCCTTAAAAGGAATAACCGCCCTCCGGTTGTTTACCCCCTCCCCCTGGA
>NV54 (*Streptomyces* spp., 95 %)

TTAGTGGCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGATAAGCCCTGGAAACGGGGTCTAA
TACCGGATATGAGCCTGGGAGGCATCTCCCGGGTTGTAAAGCTCCGGCGGTGAATGATTATCCCGCGGCCTATCT
ACTTGTTGGTGAGGTAATGGCTCCCAAGGCGACCAGGGGTTTCCGGCCTGAGAGGGCGACCGCCCCCTGGGACTA
AAAACCGCCCCCACTCCTACCGGAAGCCCCATGGGGAATATTGCCCCATGGCGAAAATCCCTCTCTCCCCCCACCC
CCCCTTTAATGGAATAACCACCGCCGGTTGTTTGCCCCCCACCCCAGGG

9.1.2. Cuban actinomycete isolates, *recA* gene sequences :

>NV2 (*S. ambofaciens recA* gene, 94 %)
CTGGACGTCCCTGGTAAGACGACCTGACCCTGCTCGCGGTGGCCAACGCCCAGAAGGCCGGCGGCCAGGTCGCCT
TCGTGGACGCGGAGCACGCCCTCGACCCCGAGTACGCGCGCAGGCTCGGCGTCGACATCGACAACCTGATCCTCC
CCCAGCCCGACAACGGCGAGCAGGCCCTGGAGATCGTGGACATGCTGGTCCGCTCCGGCGCCCTCGACCTCATCG
TCATCGACTCCGTCGCCGCGCTCGTCCCGCGCGCGGAGATCGAGGGCGAGATGGGCGACAGCCACGTCGGTCTCC
AGGCCCGTCTGATGAGCCAGGCGCTGCGGAAGATCACCAAGGCGTTCATACAGTTCAAGACTCCGTGATCTTCCT
TTTTTCCCTTCCCTCAAATAATATC
>NV3 (*S. lividans recA* gene, 94 %)
TCTACGTCTCTGGTAAGACGACCTGACCCTGCATGCGGTGGTGAACGCGCAGAAGGCCGGCGGCCAGGTCGCCTT
CGTGGACGCCGAGCACGCCCTCGACCCCGAGTACGCGAAGAAGCTCGGAGTCGACGTCGACAACCTGATCCTCTC
CCAGACCGACAACGGCGAGCAGGCCCTGGAGATCGTGGACATGCTCGTCCGCTCCGGCGCCCTCGACCTCTTCGT
CGTCCACTCCGTCGCCGCGCTCGTCCCGCGCGCGGACATCGAGGGCGAGATGGGTGACTCCTTTTCGGTCTCCAAG
CCCCTCTGATTCACCCCGGGCGCCACGGGAAATATTCCCAGTAGCGTTCCCCCCTTCCTATGGCCCCCCAAATCTT
TCTTCTTTCCTCTCCCCCAGAAA
>NV5 (*S. ambofaciens recA* gene, 93 %)
CCCGTCGCTGGTAAGACGACCTGACCCTGCGCGCGGTGGCGAACGCGCACAAGGCCGGCGGCCAGGTGGCCTTCG
TGGACGCCGAGCACGCCCTCGACCCCGAGTACGCGAAGAAGCTCGGCGTCGACCTCGACAACCTGATCCTCTCCC
AGCCCGACAACGGGGTAGCAGGCCCTGGAGATCGTGGACATGCTCGTCCGCTCCGGCGCCCTCGACCTCATCGTC
ATCGACTCCGTCGCCGCGCTCGTCCCGCGCGCGGAGATCGAGGGCGATATGGGTGACAGCCACGTCGGTCTCCAT
GGCCGCCTGATGACCCCGGGGCTGCCGAAAATCCCATTACGCATCTTTCGTTCCTGAAGACCTGGCATTTTCATTC
CTTCCGCATCCCTTAAAAAAATTC
>NV6 (*S. lividans recA gene*, 92 %)
ACGCGTTCCTTCGGTAAGACCAGCCTGACCCTGCTCGCGGTGGCGAACGCGCAGAAGTCCGGCGGCCAGGTCGCC
TTCGTCGACGCGGAGCACGCCCTCGACCCCGAGTACGCGAAGAAGCTCGGCGTCGACATCGACAACCTGATCCTG
TCCCAGCCGGACAACGGCGAGCAGGCCCTGGAAATCGTGGACATTCTGGTCCGCTCCGGTGCCCTCCACCTCATTC
GTCATTCGACTCCTTCGCCGCGCTCGTTTGCCTTCTCGACAAAATCGAGGGGGAAAATGGGGGAATCGCCCCTTTC
GTCTGCCTCGGCCGCCTAAATTAACCCCCGACCCCCGGAATTACTATCTTATTGCTGCCCTCCCACACCTTACCTTT
GTGAATGTGTTTCTCCCCTCA
>NV9 (*S. ambofaciens recA* gene, 84 %)
TTTGCATTGCCTTGAAAGACGACGCCGAGCCTGCACGCCGTGGCCACACGCGCGGAAGGCCGGCGGACAGGTCGC
ATTCGTTGACGCGGAGCACGCCCTCTACCCCGAGTACGCAAAGAAGCTCGGCGTCTACATCGACTACCTGATTCTC
TCTCCGCCCGACTACAGCAAGCTGGCCCTGGAAATCATGGACATGTTGGTCCGGGCCGGTGCCCTCTACTCATCCT
CCTCCACTCCATCCCCGCCCTCGTCTCGCGCGCGGAAATCGAGGGCGATATGGGCTACTCGCGCGTCGGTCTGCTC
GCTCGACTGATGACGCCCGCCCTGCGAGAAAACTCTTTGGGCTCGTCCAGTCCCAGACCACAGAGATCTTCTTCAT
CCGGCTCCGCAGACAAGAATG
>NV13 (*S. lividans recA* gene, 89 %)
TCTCCGTGCGTTTTCAAGACCAGCCTGAAATGCTTACGGTGGGAAGCGCCCAAAAAGCCGGCGGCCTCGTCGCCT
TCGTGGACGTCGAGCACGCCCTCGACCCCGAGTACGCAAAGAAGCTCGGCGTCGACATCGACACCTTGATCATCT
CCCAGCCGGACAACGGCGACCGGGACCTGGAGATCGTGGACATGCTCTTCCGCTCCGGCGCCCTTGACCTCTGCG
TCATCGACTCCGTCGCCGCCCTCGTCCCGCGCGCGGAGATCGAGGGCTAGATGGGCGACAGGCTCCTCCGTATCC
AGGCCCGTCTGATGATCCGGCCCTGCCGGAAGATTCGCCTTAACCTCCTTCCTTTCCCATACCCCCCTGAAGCTCCT
CCTTCCCCCCCCCCCAAAAAATGA
>NV18 (*S. lividans recA* gene, 89 %)
CCCCCGTGCTTGGTAAGACGAGCCTGAACCTGCCTGCGGTGGCGAACGCGCAGAAGGCAGGCGGCCAGGTCGCCT
TCGTGGACGCGGAGCACGCCCTCGACCCCGAGTACGCAAAGAAGCTCGGCGTCGACATCGACTACCTGATCCTCT
CCCAGCCGGACAACGGCGATCAGGCCCTGGAGATCGTGGACATGCTGGTCCGCTCCGGCGCGCTCGACCTCGTCG
TCATCGAGTCCGTCCCCGCGCTCGTCCCCCGGGAAGACATCTAGGGCAATATGGGCTACGGCCCCATGGGTCTGC
CAGGCCCGGGAGATTAGCCTTGTCCCTGTCGGAAGATCCCCGTCGTCACACTCCTTATCCTGACCCCCGGTAAATT
TGCCCCCCCCCCCTCGTGCCAAAAA
>NV34 (*S. lividans recA* gene, 90 %)
TTGCGTGGCTTTTATAAGACCGGCCTGACGCTGCTCGCGGTGTGCGAACGCGCAGAAGGCAGGTGGCCAGGTCGC
CTTCGTGGACGCGGAGCACGCCCTCGACCCCGAGTACGCCAAGAAGCTCGGCGTCAACATCGACAACCTGATCCT
CTCCCAACCGGACAACGGCGATCAGGCCCTGGATATCGTGGACGTGCTGGTCCGCTCAAGTGCCCTGGACCTCTTC
GTCATCGACTCCGTCGCCGTCCTCGTCCCGCTCGCGGATATCCACGGTCACATAGGCCTACAGCCTTTTTGTCTCTC

TGCCGGTCTACTTACCCCGCCCCGCATCGAATATCCACTCTCCCCCTTCCCTCTATATGTCCGCCTATCCCTTCCTTC
CCCCCTAAAAAAATCCTTT
>NV35 (*S. lividans recA* gene, 90 %)
TCCCCGGCCCTGGTAAGACGACCTGACGCTGCAGGCGGTGGGGAACGCGCACAAGGCAGGTGGCCAGGTCGCCTT
CGTGGACGCGGAGCAATCCCTCGACCCCGAGTACGCTAAGAAGCTCGGCGTCGACATCGACAACCTGATCCTCTC
CCAGCCGGACTACGGCGAGCAGGCCCTGGAGATCGTGGACATGCTGGTCCGCTCCGGCGCCCCGGACCTCCTCGT
CATCGACTCCGTCCCCGCCGCTCCTCCCGCGCGCCGATATCGACGGCAAAATTGGGTACCGCTCTTCCGTCTCCCA
TGCCCGTTTATTAACCCCCGCCCAGACGGAAAAATTCTCTCTGGCCCCTCTACTCCTTAGAAAACCTCCTAATTTTT
TTTCAGCCTCTCCCTCGAGAAA
>NV43 (*S. lividans recA* gene, 94 %)
CCGTTGTGCTCCGTAAGACCAGCCTCACCCTGCTCGCGGTGGCGAACGCTCAGAAGGCCGGCGGACAGGTCGCCT
TCGTGGACGCCGAGCACGCCCTCGACCCCGAGTACGCGCGCAAGCTGGGCGTCGACATCGACAACCTGATCCTCT
CCCAGCCGGACAACGGCGAACAGGCGCTGGAGATCGTGGACATGCTCGTCCGCTCCGGCGCGCTCGACCTCATCG
TCATCGACTCCGTCGCCGCGCTCGTCCCGCGGGCGGAGATCGAGGGCGAGATGGGCGACAGCCACGTGGGTCTGC
AGGCCCGGCTGATGAGCCAGGCCCTGCGGAAAGATCACCAGCGCGCTCAACCAGTCCAAGACCACGGCGATCTTC
ATCGGCCAGCTGCCCCAAAAAAATC
>NV54 (*S. lividans recA* gene, 92 %)
CTCGTCCTTGGTGAGACGAGCCTGACCCTGCTCGCCGTGGCGAACGCGCAAAAGGCCGGCGGCCAGGTCGCCTTC
GTGGACGCCGAGCACGCCCTCGATCCCGAGTTGGCGCGCAAGCTCGGCGTCGACATCGACTACCTGATCCTCTCC
CAGCCCGACAACGGCGAGCAGGCCCTGGAAATCGTGGACATGCTGGTCCGCTCCGGCGCCCCTCGACCTCATCGTC
CTCGACTCCGTCGCCGCGCTCGTCCCCCGTGCGGAGATCCAGGGCGAGATGGGCGACATCACATGGGTCTCCTTG
CCCTCTGATTAGCCAGGCCTGCGCCAGATCCCCACCAGCTCTCTCCCGTTCCCAGAGCCCCCTCAACTTCCTCTCCT
CCTCCCTCAAAAAATTGTCTTCT

9.1.3. Cuban actinomycete isolates, KSα gene sequences :

>NV9 (*S. halstedii*, sch, 88 %)
TTCCGCGGGTTTTCAAGGCCTTTACCCCGAAGGCCCCTGCCGGTTCTTCTACATCCACCGTGGCGTTCCTTCTAGCA
TGGGGGAGGGCGGCGCGGTGCTGGTCCTGGAGGAACTGGAGCACCCCCGCGCCCGCGGTGCGGATGTGTTCTGCC
AACTCTGTGGCTACTCGCCCTTCGGTGACGCTCTACCTCTTTGAGTCGACGGATCCGCTCAGGGCCTGGATATGGC
CCCGGCAACCCACACCGCGCGCCATGCTCGTCCCCTTCGAGTCCAACTGATCCGGTCCGTTCAACGCCCCACGGTC
CGGGTCCCGGCTACAACTACGTTCACCAGATCGTCGCGGTTAAATCCGTGCCCGGGCCCCCCCCCCTATCGAACCC
CGCCTTGCATTTCTCCTAAC
>NV19 (*S. albus* KSα, 84 %)
TTTCCTTGTGTTACAACGCCTACCCCGCCCACGCCTCGCGCCCCTTCGACGCGAGCCGCAACGGCTTCGTCCTCGG
CGAGGGCAGCGCCGTGTTCGTGCTGGAGGACATGGACGCCGCCCGGGCCCCGCGGGGCCCACGTCTACACCGATAT
CGCCGGCTACGCGTCCCGCAGCAACGCGTACCACATGACGGGACTGCGGCCCGACGGCACGGAGATGGCCGAGG
CGATCCGCGCCGCGCTCCACGAGGCCGCCGTCCGCCCCGACAGCGTCGACTACGTCAACGCGCACGGCTCCGGGA
CCCGGCATAACGACCGGCACGAGACGGCCGCCTTCAAGAAGACCTCCGTTGACCACGCGTTCCGGGTGCCCGTCA
GCTCGATCAAATCCATGTTGGGGCT
>NV22 (*S. venezuelae, jad*, 86 %
TTCCGTTTTTAAACCCTGCGCCACGAGAACCCCGCCCACGCCTCGCGCCCCTTCGACCCGAGCCGCAACGGCTTCG
TCCTCGGCGAGGGCAGCGCCGTGTTCGTGCTGGAGGACATGGACGCCGCCCGGGCCCGCGGGGCCCACGTCTACG
CCGAGATCGCCGGCTACGCGTCCCGCAGCAACGCGTACCACATGACGGGACTGCGGCCCGACGGCACGGAGATG
GCCGAGGCGATCGGCGCCGCGCTCCACGAGGCCGCCGTCCGCCCCGACAGCGTCGACTACGTCAACGCGCACGGC
TCCGGGACCCGGCAGAACGACCGGCACGAGACGGCCGCCTTCAAGAAGAACCTCGGTGACCACGCGTTACCGGG
TTGCCCGTCAGCTCGATCAATTCCAT
>NV49 (*S. argillaceus, mtm*, 90 %
TTTGACTTTGGACCTTGTTTCACGACCCCGCCCACGCCTCGAGCCCCTTGGGCGTTTGCCAAATCACGGGGTTCGT
CCTCGGGGAGGGCAGCGCCGTGTTCGTGACTGGTAGGTACATGGGACGACCGCCCGGGTCCCGCGGGGTTCCACG
TGGTACGTCCTAGTATCGCCGGCCTACGTCGTCCCGTCAGCAACGCGGTACCACATGACGGGACTGGGGATCCGA
CGGGACGGAGTATGGCCGAGGCGATCGGCGGCCGTGCTCCACGAGGGCCGCCGTCCGCCCCGTACAGCGTTGTAC
TACGTCAACGCGCACGGTTTCGGGGACCCGGTAGCAACGACCGTTACGAGACGGTCGTGTTGAAGAAGAGCCTTG
GTGACCACGCGTTACCGGGTGCCC

9.2. Soil clone sequences

9.2.1. Bulk soil, KSα

>B1 (*Kibdelosporangium aridum, ard*, 80 %)
TTCGACGCCATCAAGGCCACCACCCCGCGCAACGACCAACCCGAGACCGCCTCCCGTCCCTTCGACGGGTCACGC
AACGGCTTCGTGCTGGGCGAGGGCTCGGCGATGTTCGTGCTCGAGGAACTGGAGGCGGCGCGGGCACGCGGCGC
CCATATCTACGCAGAGATCGCCGGTTTCGGCTCCCGCTGTAACGCCTTTCACATGACCGGGCTCCGGCCCGACGGC
GCCGAAATGGCCGCCGCGATCACGGCCGCCCTGGACGAGGCGAAACTGAACCCGCAGGACATCGACTACATCAA
CGCCCATGGCTCGGGAACCAAGCAGAATGACCGGCACGAGACCGTCGCTTTCAAACGCAGCCTCGGCGACCATGC
GTATCACGTTCCGGTCAGTTTCCATC
>B2 (*S. arenae, ncn*, 83 %)

TTCGACGCCATCAAGGCCATTTTTACGTACAACGACGACCCGGAACACGCCGCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCACGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTTTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACCGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAAGCGGAGCCTCGGGGAAGCA
CGCCTACCGACACCGGTCAAGCTCGAT
>B3 (*S. arenae, ncn*, 83 %)
TTCGACTTCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCG
CAACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGC
GCGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACACCTACCACATGACCGGCCTCAAGCCCGACGG
CCGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTTGGACGTTGGACTACGTCA
ACGCACACGGGTTCGGGCACCAAGCAGAACGGCCGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCAC
GCCTACCGCACACCGGTCAGCTCGAT
>B4 (*S. arenae, ncn*, 84 %)
TTCGACGCCATCAAGGCCACTATTACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCTCGACGGC
CGGGAGATGACCGAGGCGGTCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGACACCGGTCAGCTCGATCAA
>B5 (*S. venezuelae, jad*, 84 %)
TTCGACTTCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B6 (*S. venezuelae, jad*, 84 %)
TTCGACTTCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B7 (*S. arenae, ncn*, 83 %)
TTCGACGCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGACGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B8 (*S. arenae, ncn*, 83 %)
TTCGACGCCATCAAGGTCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGGCACACGGGTCGGGGCACCAAGCAGAACGACCGGCACGAGAACGGCGGGTTCAAACCGGAGCCTTGTAAGCA
CGCCTTTCGGAAATCTGCAAGCTTATC
>B9 (*S. venezuelae, jad*, 80 %)
TTCGACGCCATCAAGGCCACCTCGGCCCGCAACGACGACCCGGAGCATGCGTCGCGGCCCTTCGACCGCGACCGC
GACGGGTTCGTGATGGGAGAGGGCTCAGCGGTCTTCGTCATCGAGGAGCCGGACCGGGCGATGGCCCTGCGGCGC
GCACATCTACTGCGAGGTGGCCGGCTACGCGAGCCGGGGCAACGCGTTCCACATGACCGGGCTGCGGCCCGACGG
CCTCGAGATGGCCGAGGCGATCACCGATGCGATGGCGCAGGCCCGCTACGTGTCGGAGGACGTCGACTACATCAA
CGCGCACGGCTCGGGCACGAAGCAGAACGACCGCCACGAGACCGGCGCGTACAAGCGGGTGCTCGGCGATCACG
CCTACAAGACGCCCGTGAGCTCCATCA
>B10 (*S. venezuelae, jad*, 84 %)
TTCGATTTCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGTCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B12 (*S. arenae, ncn*, 84 %)
TTCGACGCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B13 (*S. arenae, ncn*, 84 %)

TTCGACGCCATCAAGGCCATTTTTACGTACAACGACGACCCGGAACACGCCGGCCCGGCCCTTCGACGGCACCCG
CAACGGCTTCGTACTGGCCGAGGGTGCGGCCATGGTCGTCCTGGAGGAGCTGACGGCCGCCCGGCGCCGGGGCGC
GCGCGTCCACGCCGAGGTGGGCGGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACG
GCCGGGAGATGGCCGAGGCGATCCGCGCGGGGTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGACA
ACGCACACGGGTCGGGCACCAAGCAGAACGACCGCCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCAC
ACCTACCGCACACCGGCAGCTCGACAA
>B14 (*S. venezuelae, jad*, 80 %)
TTCGACGCCATCAAGGCCATTTTGACGTACAACGACGACCCGGAACACGCCTGCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCCTGCTGGAGGAGCTGACGGCGTCCGGCGCCGGGGCGCCC
GCGTATACGCCGACGTGGGCGGGTATGCCACCCGCTGCAACGTCTACCACATGACCGGCTCAAGCCCGACGGCCG
GGAGAGGACCGAGGCGATCCGCGCGGAGGTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAACG
CACACGGGTCGGGCACCAAGCAGAACGACCGGCACGAAACCCGCGGCGTTCAAGCGGACCCTCGGTGAGCACGC
CTACCGCACACCGGTCAGCTCGATCAA
>B16 (*S. arenae, ncn*, 84 %)
TTCGACGCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>B17 (*Saccharopolyspora hirstu, hir*, 86 %)
TTCGACGCCATCAAGGCCACCACACCCAACAACGACGCCAACGCCGCCGGCACCTCCAGGCCCTTCGACGGCCGC
CGCGACGGCTTCGTCCTCGGCGAGGGCGCCGCCGTCATGGAGATCGAGGAGGCAGCAGCGGCCCCGCCGCCGCGG
AGCCCGGACCTACGCCGAGGTGGCCGGCTTCGCCAGCCGCTCCAACGCCTACCACATGACCGGGCTCAAGCCCGA
CGGCCGCGAGATGGCCGAGGCCATCGGCGTCGCCCTGGGCCCGGGCCCCGCCTCGATGCCACGGACATCGGCTACAT
CAACGCGCACGGCTCGGGCACCAAGCAGAACGACCGCACGAGACGGCCGACTTCAAACGCAGCCTGGGACAGCA
CGCCTACACATCCCGGCACTCAGCAAG
>B20 (*Saccharopolyspora hirstu, hir*, 83%)
TTCGACGCCATCAAGGCCACCACTCCCCGTAACGACGATCCGGAGCACGCCTCCCGGCCCTTCGACGCGAGCCGC
AACGGGTTCGTACTTGGCGAGGGCGCGGCGGTGTTCGTACTGGAGGAGCTCGGGCACGCCCGGCGACGCGGCGC
GCGGATCTACGCCGAGATCGCCGGGTTCGCCTCCCGGTGCAACGCCTTCCACATGACCGGACTCCGCCCCGACGG
ACGCGAGATGGCCGAGGCGATCCGGGTGGCGCTCAACGAGGCGCGGATCGATCCCACCGAGATCGACTACATCA
ACGCCCACGGATCGGGCACCCGGCAGAACGATCGGCACGAGACGGCGGGCTTCAAACGCAGTCTCGGCGATCAC
GCTTACCGGACCCCGGTAGTCGATCAGT

9.2.2. Bulk soil, KSβ

>B1 (*S. cyanogenus, land*, 82 %)
GTCGGCGAAGACGACTTTGACCTCACCGGGCGTCACGCCGGCATCGGCCAGCGCGCCCACGATGGCGCGCTTGAG
CGCCGGCTCCCGCCCGCTGCCGGGCCGGGGATCGAAGGTGGCGGCGTACCCGGCGACCTCGCCGTACACCGTCGG
CGCCCCCCGCTCCCGTGCGGCCTGGGCGTCCTCTAGGACCAGGATCGCTCCGGCCTGGGCCGGTACGTGCCCTGCC
GCCCGGTCATTGAAGGGAAGGTAGGCCGCCGCGGGGTCACCATTCTCGCTGAGCCGCCCGCCGGCCATTCTGGCC
GACCCAGCCCCAGGGGCAGATGGCCCCGTTCAGACACGCCGCGCAAGGATGAAGGTTGCTCCCCCGTGGATGTGC
CCCCGGCCTGGGCGAATCGCGTCA
>B2 (*S. roseofulvus,fren*, 90 %)
GTCGGCGAAGACGACGTCGGATCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGC
AGCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACG
GGCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCT
CGCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGG
AGACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTG
CGCCGTGCCTGGCCGAGCGCGTCCAGCCC
>B3 (*S. roseofulvus,fren*, 90 %)
GTCGGCGAAGACGACGTTTTACCTCGACGGCGGCGGTCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
ACCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAAGTGGGAG
ACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGTCGGCACCTGAGGGGCACGCCCGCCCGGATTGCTTGCGC
CGTGCCTGGGCGAGCAGCCTGCAGCCC
>B4 (*S. roseofulvus,fren*, 90 %)
CTTGGCGAATACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCAG
CCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGGG
CGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTCG
CCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGAG
ACCCAGCCCCAGGGGTCGAAGGCGGAGTTCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGCG
CCGTGCCTGGCCGAGCGCGTCCAGCCCG
>B6 (*S. roseofulvus,fren*, 92 %)
GTCGGCGAAGACGACTTTGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCAG
CCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGGG
CGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGTCCTCACCGGGGACGTAACCGCTCG

CCCCGGTGTCTAAGGGCCCGGTAGGCGGTTTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGAG
ACCCACCCCCAGGGGTCGGAATGCTGAGCCCACAGCACCGGTCACCGATGAGGGGCACTCCCGCCTCGGATGCTG
GAGCCGTGCCTGGCCGAACCGCGTTC
>B8 (*S. roseofulvus, fren*, 90 %)
GTCGGCTTAAGACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGGCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGC
GCCGTGCCTGGCCGAGCGCGTTCAGCCC
>B9 (*S. venezuelae, methlpik*, 90 %)
GTCGGCGAAGACGACGTCGACGTCGGATGGACCCACGCCGGCCCGCTCGAGCGCCTGGCTGATCGCGCGGTCGAG
GTGGCGGCCGCTGGGGTCGCAGTCCGCGTAGGGGTAGGCGTCGAACGTCGCCGCGTAGCCGGCGACCTCGCCGTA
CACCTGCGGGGCGCCGCGCTCCCTCGCCGAGGCCAGCTCCTCGACCAACACGAGGGCGCCGCCCTCGCCGGGCAC
GTAGCCGCCGGCGCGCGCGTCGAACGGCCGGTAGCCGTTCGCCGGATCGGTCGCCTCGCTCACCTGACGGGTCGT
CGTCTGGCACACGGTCCCGTAGGGCGCGACCGGCGCTTCGGTGCCACCGGCCACCACAGACGTGAACCCGCGCCG
ATCTCCTTGTGGCCTGCAGAGCGTT
>B11 (*S. glaucescens, tcm*, 84 %)
GTCGGCGAAGACGACGTCGACGGCGTCGGCGCTCAGGTGCGCGTCGGCGAGGGCGCGGCGGATCACCGAGGCGA
GGACCGGGGGGCGGCCCGAGCCGGGTGCCGGGTCGAAGCCGGCGGCGTGGCCGAGGATGCGGCCGTAGTGGCCC
ACCCCGCGCTCACGGGCACTGTCCTCGGTCTCGACGACCATGATGGCTCCGCCCTCGCCGGGGACGTAGCCGCGC
GCCGCCTCGTCGAAGGGCAGTAAGGCGCGGGTCGGGCCGTCCTCGGTGGACACGTGGCCGGTGGACAGCTGGGC
GGTCAGCCCGTAGGGGCACAAGGAGGCGTTGTGCCGCCGGAGAGCACCAACGAAGCCGGTGTCAACATCTGCTG
GACTGCCCAAGGCTCCAGCGCGCCCGCCG
>B12 (*Saccharopolyspora hirsuta, hir*, 86 %)
GTCGGCGAAGACGACGTCTATCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCAG
CCCGGGCGGTCGCCCGGAGTCCGGGGCCGGGTTTAATTTGGCGGAGTACCCGGCGAGTTCGCGTACACACGGGCG
CCCCGGTCGTCCGCCGACCGGGCGTTCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTCGCC
CCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGAGAC
CCACCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGCGCCG
TGCCTGGCCGAGCGCGTCAAACCGCCG
>B13 (*S. roseofulvus, fren*, 87 %)
GTCGGCGAAGACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTGGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGC
GCCGTGCCTGGCCGAGCGCGTCCAGCCGC
>B14 (*S. roseofulvus, fren*, 90 %)
GTCGGCGAAGACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCTGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGC
GCCGTGCCTGGCCGAGCGCGTCCAGCCCG
>B19 (*S. roseofulvus, fren*, 83 %)
GGCGGCGAAGACGACGTCGACGTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCTCAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGACGGTCACCATGAGGGGCACGCCCGCCCGATGCTGCG
CCGTGCCTGGCCGAGCGCGACAGCCCTCG

9.2.3. Rhizosphere soil, KSα

>R3 (*S. halstedii, sch*, 85 %)
TTCGACGCCATCAAGGCCACCTCGCCGAACAACGACGACCCGGAGCACGCCTCTCGACCGTTCGACGCCAACCGG
GACGGCTTCGTGATGGGCGAGGGCGGCGCGGTGCTCGTACTCGAAGAGCTCGAACACGCCCGCGCCCGCGGCGC
GACCGTCTACTGCGAAATCCAGGGCTACGCCACCTTCGGCAACGCCTACCACATGACCGGGCTGACCGCGGACGG
ACTCGAGATGGCCCAGGCCATCAACTCCGCCCTCGACCACGCCCGCCTGGATGCCACCCGGATCGACTACGTCAA
CGCGCACGGCTCGGGTACCAAACAGAACGATCGGCACGAGACGGCGGCGGTGAAGCGGTCCCTCGGCACACACG
CCTACCGGATACCGGATGAGTTCGATC
>R4 (*S. halstedii, sch*, 79 %)
GCCGACCCCATCAAGGCCACCTCGAACGAACAACGACGACCCGGAGCACGCCTCCCGACCGGCCGACGCCAACC
GGGACGGCATCGTGAAGGGCGAGGGCGGCGCGGCGCCCGTACACGAAGAGCACGAACACGCCCGCGCCCGCAGC
GCGACCGACCACGGCGAAAGCCAGGGCTACGCCACCATCGGCAACGCCTACCACATGACCGGGCAGACCGCGGA
CGGACACGAGAAGGCCCAGGCCATCAACACCGCCCACGACCACGCCCGCCTGGACGCCACCCGGAACGACTACG
CCAACGCGCACGGCACGGGGACCAAACAGAACGACCGGACGAGACGGCGGCGGAGAAGCGGCCCCACGGCACAC
ACGCCCACCGGAACCGACGAGTACAACAAGC

>R5 (*S. halstedii, sch*, 84 %)
TTCGACGCCATCAAGGCCACCTCGCCGAACAACGACGACCCGGAGCACGCCTCTCGACCGTCCGACGCCAACCGG
GACGGCTTCGTGATGGGCGAGGGCGGCGCGGTGCTCGTACTCGAAGAGCTCGAACACGCCCGCGCCCGCGGCGC
GACCGTCTACTGCGAAATCCAGGGCTACGCCACCTTCGGCAACGCCTACCACATGACCGGGCTGACCGCGGACGG
ACTCGAGATGGCCCAGGCCATCAACTCCGCCCTCGACCACGCCCGCCTGGATGCCACCCGGATCGACTACGCCAA
CGCGCACGGCTCGGGTACCAAACAGAACGATCGGCACGAGACGGCGGCGGTGAAGCGGTCCCTCGGCACACACG
CCTACCGGATCCGATGAGTTCGACAAG
>R7 (*S. halstedii, sch*, 83 %)
TTCGACGCCATCAAGGCCACTCGCCGAACAACGACGACCCGGAGCACGCCTCTCGACCGTTCGACGCCAACCGGG
ACGGCTTCGTGATGGGCGAGGGCGGCGCGGTGCTCGTACTCGAAGAGCTCGAACACGCCCGCGCCCGCGGCGCG
ACCGTCTACTGCGAAATCCAGGGCTACGCCACCTTCGGCAACGCCTACCACATGACCGGGCTGACCGCGGACGGA
CTCGAGATGGCCCAGGCCATCAACTCCGCCCTCGACCACGCCCGCCTGGATGCCACCCGGATCGACTACGTCAAC
GCGCACGGCTCGGGTACCAAACAGAACGATCGGCACAGAGACAGCGGACGGTGAAGCGGTCCCTCGGCACACAC
GCCTACCGGATACCGATGAGTTCGATC
>R9 (*S. cinnemonensis, mon*, 83 %)
TTCGACGCCATCAAGGCCCCCCACACCCAGCAACGACGACCCCCAGCACGCCTCACGCCCCTTCGACCTCGACCG
CAAGGGCTTCGTCCTCGGCGAAGGCTCCGCCGTACTCGTCCTCGAAGAACGCACCGCGGCCCTACGACGCGGAGC
ACACATCTACGCGGAGATAGTGGGATTCGCCAGCCGCAGCAACGCCTACCACATGACCGGCCTCAAACCCGACGG
CCGCGAAATGGCCGAAGCCATCCGAGCCGCATGGACCAAGCCCGCCTCAACCCCGAAGACGTCGACTACATCAAC
GCCCACGGCTCCGGAACCCGGCAAAACGACCGCCACGAAACCGCCGCCTTCAAAAACGCACTCGGCGACCACGC
CTACACCATCCCGGTCAGCTCCATCA
>R11 (*S. arenae, ncn*, 84 %)
TTCGACGCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGACGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>R12 (*Sachharopolyspora hirsuta, hir*, 88 %)
TTCGACGCCATCAAGGCCATTCCACCCAACAACGACGCCAACGCCGCCGGCGGCTCCAGGCCCTTCGACGGCCGC
CGCGACGGCTTCGTCCTCGGCGAGGGCGCCGCCGTCATGGTGATCGAGGAGGCAGCAGCGGCCCCGCCGCCGCGG
AGCCCGGATCTACGCCGAGGTGGCCGGCTTCGCCAGCCGCTCCAACGCCTACCACATGACCGGGCTCAAGGCCGA
CGGCCGCGAGATGGCCGAGGCCATCGGCGTCGCCTGGGCCGGGCCCGCTCGATGCCACGGACATCGGCTACATCA
ACGCGCACGGCTCGGGCACCAAGCAGAACGACCGCACGAGACGGCCGCTTCAAACGCAAGCTGGGACAGCACGC
CTACCACATTCCGGGCAACTTCATCAA
>R13 (*S. venezuelae, jad*, 80 %)
TTCGACGCCATCAAGGCCACCTCGGCCCGCAACGACGACCCGGAGCATGCGTCGCGGCCCTTCGACCGCGACCGC
GACGGGTTCGTGATGGGAGAGGGCTCAGCGGTCTTCGTCATCGAGGAGCCGGACCGGGCGATGGCCCGCGGCGC
GCACATCTACTGCGAGGTGGCCGGCTACGCGAGCCGGGGCAACGCGTTCCACATGACCGGGCTGCGGCCCGACGG
CCTCGAGATGGCCGAGGCGATCACCGATGCGATGGCGCAGGCCCGCTACGTGTCGGAGGACGTCGACTACATCAA
CGCGCACGGCTCGGGCACGAAGCAGAACGACCGCCACGAGACCGGCGCGTACAAGCGGGTGCTCGGCGATCACG
CCTACAAGACGCCCGTGAGCTCCATCA
>R16 (*S. arenae, ncn*, 84 %)
TTCGACGCCATCAAGGCCACCTCGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA
>R17 (*S. halstedii, sch*, 83 %)
TTCGACGCCATCAAGGCCATTTTGCCGAACAACGACGACCCGGAGCACGCCTGCCGACCGTTCGACGCCAACCGG
GACGGCTTCGTGATGGGCGAGGGCGGCGCGGTGCTCGTACTCGAAGAGCTCGAACACGCCCGCGCCCGCGGCGC
GACCGTCTACTGCGAAAATCCAGGGCTACGCCACCTTCGGCAACGCCTACCACATGACCGGGCTGACCGCGGACG
GGACTCGAGATGGCCCAGGCCATCAACTCCGCCCTCGACCACGCCCGCTGGATGCCACCCGGATCGACTACGTCA
ACGCGCACGGCTCGGTACCAAACAGAACGATCGGCACGAGACGGCGGCGGGGAAACGGTCCTTGGGACACACGC
CTACCGGATACCGATCAGTCGATCAAA
>R20 (S. arenae, ncn, 84 %)
TTCGACGCCATCAAGGCCACCACGACGTACAACGACGACCCGGAACACGCCTCCCGGCCCTTCGACGGCACCCGC
AACGGCTTCGTACTGGCCGAGGGTGCGGCCATGTTCGTCCTGGAGGAGCTGACGGCCGCCCCGGCGCCGGGGCGCC
CGCGTCTACGCCGAGGTGGGCGGCTATGCCACCCGCTGCAACGCCTACCACATGACCGGCCTCAAGCCCGACGGC
CGGGAGATGACCGAGGCGATCCGCGCGGCGCTCGACCAGGCCAGGCTGGACCCCTCGGACGTGGACTACGTCAA
CGCACACGGCTCGGGCACCAAGCAGAACGACCGGCACGAGACCGCGGCGTTCAAGCGGAGCCTCGGCGAGCACG
CCTACCGCACACCGGTCAGCTCGATCA

9.2.4. Rhizosphere soil, KSβ

>R2 (*S. roseofulvus, fren*, 87 %)
GTCGGCGAAGACGACGTCGAATCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGC
AGCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACG

GGCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCT
CGCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGG
AGACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTG
CGCCGTGCCTGGCCGAGCGCGTCCAGCCC
>R3 (*S. coelicolor, whiE*, 83 %)
GTCGGCGAAGACCACGTCGACCTCGACGGGCCCGCAGCGCGCCGCTTCGCAAGGCGCCACGGATCGCGTGCGCCA
GTCCCTCCCGGGACTCCTCCCACCGGGACGCCCGGGAAAAGGTGGTGGCAGTGCCCCGGCACAGCTTGGTCCTGA
TGGCGCGCCCCGGCTCCCTGATCGTGGTCCTTTGGTCCTCAACTCGAAGAACATGGAGCCCCCCCTAGGCCGGCAC
GAACCCGCAGGTGCCCAATGCGAAAAGGCCCGTTCTGGCACGCCTAGGGCCCGCCTGCGCTTACCCGCGCTTTGG
ACCCGTACGTGCACGGGCCTCCCAAATACGCGAACATCCAGGCCGCCGGGGAGCGTGAGGAAGCAGGCTGCGAG
GAGGCCACCTCCAGCTTGACCCGGC
>R4 (*S. roseofulvus, fren*, 87 %)
GTCGGCGAAGACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCCGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTGGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGC
GCCGTGCCTGGCCGAGCGCGTCCAGCCCG
>R5 (*S. coelicolor, whiE*, 82 %)
GTCGGCGAAGACGACTTCGACCTCGTCGGGCCCGCAGCACGCCGCTCGCAAGGCGCCACGGATCGCGTGCGCCAG
TCCCTCCCGGGACTCCTCCCACCGGGACGCCCCGGTGAAGGTGGCGGCGTGCCCGGCGAGCTCGGCCCGGATGCG
CGCCCCGCGCCCGATCGCGGTTCTTTCGTCCTACACAGGAACATGGAGCCCCCCTCAGCCGGCAACGAACCCCGC
AGGCAGCCCTATGTGAAAGGCCCGGTAGGCACGGCTGGGGCTCGTCGTTCCCGGACTTAGGCCCTTTTACCAGGG
CAGGCACAGCACCCAATACGGGGGCTTCGAAGGCCTCCGAGGGAATAAGGGACCGCCGGCTAGGATGCCATTCC
GGGATGGACACCGCGCGGTCAACCAG
>R7 (*S. coelicolor, whiE*, 81 %)
GTCGGCGAAGACGACGTCTACCTCGTCGGGCCCGCAGCGCGCCGCTCGCAAGGCGCCACGGATCGCGTGCGCCAG
TCCCTCCCGGGACTCCTCCCACCGGGACGCCCCGGTGAAGGTGGTGGCAGTGCCCGGCGAGCTCGGCCCGGATGT
GCGCCCCGTCCCGATCGTGGTCCTTTCGTCCTCGACCATGAACATGGCGCCCCCCTCGGCCGGGACGAACCCGCAT
GCGCCCGAGGTGAAGGGCCGGTAGGCACGGGTGGGGTCCTCCTCTCGTACTGAGGCCCTCGTACCGAGCTGGCAC
ACCACCGAATACGGGGCCAGCGGTGCCTCTGCGGCGAGGAGGACCACCCGGCGCGGATGCCACTCCGGATGGAC
CCGGCGGCGTGCACTAGGGCGTGAA
>R9 (*S. venezuelae, fren*, 89 %)
GTCGGCGAGACGACGTCGACCTGGTCGGGGGTGGTTCTGGCGTCGGTCAGGGCTTGGGTGATGGCGTGGTGGAGG
TTGGATCTGGGGTGATGACCTGGGGTGGGGCGTGAGTCGAAGGTGGAGGTGTATACGGCGATCGCGCCGTAGGG
GCGGGCGCCTCTGGCGCCGGCGGAGCTCTTGTTCTCGACGATCATGATGGCGCCGCCTTCGCCGGGTATGTACCCG
CAGGCCCCGGTGACTAAGGGCAAGTAAGCGCTTTCTGGGTCGAAACGGGTGCTGACCATTCCCTGGAGAGGTGTG
CGACCAAGCCGCAGGGGCATACGGAGGCGACATGCCGCCTGATGACACCATGGCGGCGTTCTCGCGGATGAGGC
GCAGGCCTTGGCAGCCAGATCGAGGC
>R18 (*S. roseofulvus, fren*, 87 %)
GTCGGCGAAGACGACGTCGACCTCGTCGGCGGCGATCCCGGCGTCGGCCAGGGCGAGTTCGGCGGCGCGCCGCA
GCCCGGGCGGCCGCCTGGAGTCCGGGGCCGGGTCGAACGTGGCGGCGTACCCGGCGAGTTCGCCGTACACACGG
GCGCCCCGGTCGTCCGCCGACCGGGCGTCCTCCAGGATGCAGAGGGCGCCGCCCTCACCGGGGACGTAACCGCTC
GCCCCGGTGTCGAAGGGCCGGTAGGCGGTGTCCGGGTCGGGGTCGGCGCTGACCCGCCCGCTGGACAGGTGGGA
GACCCAGCCCCAGGGGTCGAAGGCGGAGTCCACACCGCCGGTCACCATGAGGGGCACGCCCGCGCGGATGCTGC
GCCGTGCCTGGCCGAGCGCGTCCAGCCCG