UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOTECNOLOGÍA

IDENTIFICATION OF THE CHAXAMYCIN AND CHAXALACTIN BIOSYNTHESIS GENES THROUGH GENOME MINING OF *Streptomyces leeuwenhoekii* C34 AND HETEROLOGOUS PRODUCTION OF CHAXAMYCINS IN *Streptomyces coelicolor* M1152

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN INGENIERÍA QUÍMICA Y BIOTECNOLOGÍA

JEAN FRANCO ALEJANDRO CASTRO FIGUEROA

PROFESORES GUÍAS:
JUAN ALFONSO ASENJO DE LEUZE DE LANCIZOLLE
BARBARA ANDREWS FARROW

MIEMBROS DE LA COMISIÓN:
ZIOMARA GERDTZEN HAKIM
ORIANA SALAZAR AGUIRRE
RAFAEL VICUÑA ERRÁZURIZ

SANTIAGO DE CHILE
2015

SUMMARY OF THE THESIS FOR THE DEGREE IN: Doctor in Engineering Sciences, mention Chemical Engineering and Biotechnology
BY: Jean Franco A. Castro Figueroa
DATE: December 2015
SUPERVISORS: Dr. Juan A. Asenjo and Dra. Barbara Andrews

IDENTIFICATION OF THE CHAXAMYCIN AND CHAXALACTIN BIOSYNTHESIS GENES THROUGH GENOME MINING OF *Streptomyces leeuwenhoekii* C34 AND HETEROLOGOUS PRODUCTION OF CHAXAMYCINS IN *Streptomyces coelicolor* M1152

Streptomyces leeuwenhoekii C34 is an actinomycete isolated from the Atacama Desert, Chile, that produces new antibiotics: chaxamycins A to D and chaxalactins A to C. The goal of this work was to identify the genes involved in the biosynthesis of chaxamycins and chaxalactins and produce chaxamycins in a heterologous host.

Chapter One details the development of basic procedures for growth and genetic manipulation of *S. leeuwenhoekii* C34. The results indicated that 30 °C was the optimal temperature for chaxamycin production in liquid culture, and high sporulation levels on solid medium were achieved at 37 °C. *S. leeuwenhoekii* C34 was sensitive to thiostrepton, apramycin, hygromycinB and kanamycin, but not to nalidixic acid, antibiotics used to select for genetically modified bacteria. A temperature-sensitive vector (pGM1190) and other with an insertion site recognised by the integrase system of the phage C31 (pSET152) were successfully transferred to *S. leeuwenhoekii* C34 by conjugation with the methylation-deficient strain, *E. coli* ET12567/pUZ8002. Supplementation of the conjugation medium with either 120 mM $MgCl_2$ or 60 mM $CaCl_2$, significantly increased conjugation frequency compared to no salt addition. Thus, it was demonstrated that *S. leeuwenhoekii* C34 was able to accept foreign unmethylated DNA.

Chapter Two describes the identification and heterologous expression of the chaxamycin biosynthesis gene cluster. The genome of *S. leeuwenhoekii* C34 was sequenced *de novo* by combining Illumina and PacBio technologies. A group of 27 genes that span a region of 80.2 kb (locus 1,210,347–1,290,550), encoded for the necessary enzymes for chaxamycin biosynthesis. pIJ12853 contained a 145 kb insert of *S. leeuwenhoekii* C34 genomic DNA that included the 80.2 kb sequence, which was cloned into *S. coelicolor* M1152 that does not produce chaxamycin, giving rise to strain M1650. M1650 was able to produce all chaxamycin species, confirming that the genes present in pIJ12853 encode for the chaxamycin biosynthesis route. Genes from the heterologous host could also be involved in the biosynthesis and/or export of these molecules. Deletion of the AHBA synthase gene (*cxmK*) of *S. leeuwenhoekii* C34 gave rise to strain M1653, unable to produce chaxamycins. To prove that the abolition of chaxamycin production was solely due to its inability to produce AHBA, a liquid culture of M1653 supplemented with commercial AHBA was performed and chaxamycin was shown to be produced. This is further evidence that the chaxamycin biosynthesis gene cluster was identified.

Chapter Three describes the bioinformatic identification of the chaxalactin biosynthesis gene cluster. A sequence of 80.7 kb (locus 7,146,903–7,227,608) contained 5 genes that encoded for a PKS, whose domain architecture matches with that predicted for chaxalactin A biosynthesis. A gene that encodes for a cytochrome P450 is likely to be related to a C-14 hydroxylation reaction of chaxalactin A to give rise to chaxalactin B. Two genes encoding for *O*-methyltransferases are possibly involved in the C-13 *O*-methylation step of chaxalactin B that gives rise to chaxalactin C.

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE: Doctor en Ciencias de la Ingeniería, mención Ingeniería Química y Biotecnología
POR: Jean Franco A. Castro Figueroa
FECHA: Diciembre 2015
PROFESORES GUÍAS: Dr. Juan A. Asenjo y Dra. Barbara Andrews

IDENTIFICACIÓN DE LOS GENES DE BIOSÍNTESIS DE CHAXAMICINAS Y CHAXALACTINAS MEDIANTE MINERÍA DE GENOMAS DE *Streptomyces leeuwenhoekii* C34 Y PRODUCCIÓN HETERÓLOGA DE CHAXAMICINAS EN *Streptomyces coelicolor* M1152

Streptomyces leeuwenhoekii C34 es un actinomiceto aislado del desierto de Atacama, Chile, productor de los antibióticos chaxamicinas A a D y chaxalactinas A a C. El objetivo de este trabajo fue identificar los genes involucrados en la biosíntesis de chaxamicinas y chaxalactinas y producir chaxamicinas en un huésped heterólogo.

Capítulo Uno detalla procedimientos de crecimiento y modificación genética de *S. leeuwenhoekii* C34. La temperatura óptima para producción de chaxamicinas en medio líquido fue 30 °C, y la que permitió alcanzar altos niveles de esporulación en medio sólido fue 37 °C. *S. leeuwenhoekii* C34 fue sensible a tioestreptona, apramicina, higromicina B y kanamicina, mientras que no a ácido nalidíxico, antibióticos usados para seleccionar bacterias genéticamente modificadas. Un vector sensible a temperatura (pGM1190) y otro con un sistema de integración del fago C31 (pSET152) fueron exitosamente transferidos a *S. leeuwenhoekii* C34 por conjugación con una cepa que no metila ADN, *E. coli* ET12567/pUZ8002. Suplementación del medio de conjugación con 120 mM $MgCl_2$ ó 60 mM $CaCl_2$, incrementó la frecuencia de conjugación de forma significativa, comparado con el control sin adición de sales. En este trabajo se demostró que *S. leeuwenhoekii* C34 incorpora ADN foráneo no metilado.

Capítulo Dos describe la identificación y expresión heteróloga de los genes de biosíntesis de chaxamicinas. El genoma de *S. leeuwenhoekii* C34 fue secuenciado *de novo* combinando tecnologías de secuenciación Illumina y PacBio. Un grupo de 27 genes (80,2 kb, locus 1.210.347−1.290.550), codifica enzimas para la biosíntesis de chaxamicinas. pIJ12853 contenía un inserto del genoma de *S. leeuwenhoekii* C34 de 145 kb que incluye el segmento de 80,2 kb, el que fue transferido a una cepa que no produce chaxamicinas, *S. coelicolor* M1152, resultando en la producción de chaxamicinas A−D, confirmando que los genes presentes en pIJ12853 codifican para la ruta de biosíntesis de chaxamicinas. Genes del huésped heterólogo podrían estar involucrados en biosíntesis y/o exporte de estas moléculas. La deleción del gen AHBA sintasa (*cxmK*) en *S. leeuwenhoekii* C34, dio origen a la cepa M1653 que pierde su capacidad de producir chaxamicinas. Para demostrar que la interrupción en la producción de chaxamicinas fue sólo debido a su incapacidad de producir AHBA, un cultivo líquido de M1653 suplementado con AHBA comercial permitió restablecer la producción de chaxamicinas. Esto es una prueba más de que el los genes de biosíntesis de chaxamicinas fueron identificados.

Capítulo Tres describe la identificación bioinformática de los genes de biosíntesis de chaxalactinas. Una secuencia de 80,7 kb (locus 7.146.903−7.227.608) contenía genes codificantes de 5 subunidades de una policétido sintasa, cuya arquitectura de dominios coincidía con la predicha para biosíntesis de chaxalactina A. Un gen que codifica una citocromo P450 sería responsable de la hidroxilación C-14 de chaxalactina A que da origen a chaxalactina B. Dos genes codificantes de *O*-metiltransferasas estarían involucrados en una *O*-metilación C-13 de chaxalactina B, que da origen a chaxalactina C.

LIST OF PUBLICATIONS GENERATED FROM THIS WORK

J. F. Castro, V. Razmilic, J. P. Gomez-Escribano, B. Andrews, J. A. Asenjo and M. J. Bibb. Identification and heterologous expression of the chaxamycin biosynthesis gene cluster of *Streptomyces leeuwenhoekii*.
Applied and Environmental Microbiology, 2015, volume 81, issue 17, pages 5820−5831.
DOI 10.1128/AEM.01039-15

J. P. Gomez-Escribano, **J. F. Castro**, V. Razmilic, G. Chandra, B. Andrews, J. A. Asenjo and M. J. Bibb.
The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2.
BMC Genomics, 2015, volume 16, issue 1, page 485.
DOI 10.1186/s12864-015-1652-8

To my family

”There is a certain amount of hubris in humans thinking that they can create molecules as well as nature can”

William C. Campbell, Nobel Prize in Medicine 2015

ACKNOWLEDGEMENTS

It has been almost five years since I decided to start my doctorate studies at the University of Chile. Five years of an excellent stage of my life, of arduous work and dedication in which I've had the opportunity to travel to amazing places and met extraordinary people from different places, languages and beliefs... Definitely a life-time experience that I'll treasure in my memories.

I'd like to start thanking to my supervisors, Juan A. Asenjo and Barbara Andrews for their unconditional support from the beginning of my studies, and for contributing to develop my scientific career. I appreciate the confidence granted in me to be part of the project of microbiology of the Atacama Desert that has opened my eyes to the potential and great future of our microbial resources.

Mervyn, I'll be everlastingly grateful for your kindness and support since the first moment I contacted you. Thanks for your invaluable scientific guidance, for being an inspiration. Thanks for making my stay in the UK a lifetime experience. I can't thank you enough for that.

JuanPa (aka Juan Pablo Gómez), thank you very much for your immense contribution to my research and for teaching me the tricks of *Streptomyces* genetics. Most of the results I obtained wouldn't have been even possible without your guidance, help and inexhaustible patience. Thanks for all those the good memories and friendship. I owe you many!

Vale, thank you for your unconditional help and friendship through all these years. Emilie, thanks for your input and useful discussions related to my work. Thanks to all my colleagues and everybody that integrates the Centre for Biotechnology and Bioengineering (CeBiB) of the University of Chile, for making the working place a great environment to work in.

I'm grateful for all my lab-mates of the Department of Molecular Microbiology of the John Innes Centre (JIC) in Norwich, United Kingdom, for their comradeship, friendship and unlimited help with my daily laboratory work.

Special thanks to Michael Goodfellow and Alan Bull for your always positive energy and for being such an inspiration for new generations that are stepping into the field of actinomycetes. To Marcel Jaspars and Mostafa Rateb, thank you for making yourselves always available to answer any query. Thanks a lot to all faculty members and students of the Summer School in Applied Molecular Microbiology 2014 in Dubrovnik, Croatia for the knowledge given and your useful comments on my research.

I cannot close this acknowledgements without thanking the most important, my family. Thanks mother, brother and father for always being on my side in this journey, you are the pillar of my development as a person and this achievement is dedicated to you.

Thanks to all my family for being also part of this story.

To all people that I've run into and encountered in this process and that have indirectly contributed with my career development, thanks.

I finally thank to the National Commission for Scientific and Technological Research (CONICYT) for financial support: "National Doctorate Scholarship" (#21110356, 2011–2015) and "Operating Expenses" (2013 and 2014) and for "Assistance to Events and Short Courses for Doctorate Students Grant" (2014). To the Becas Chile Programme of CONICYT: "Doctoral Internship Abroad Scholarship" (2013–2014). To the "Basal Programme" of CONICYT (#FB0001) and the Biotechnological and Biological Sciences Research Council (BBSRC, UK) Institute Strategic Programme Grant "Understanding and Exploiting Plant and Microbial Secondary Metabolism" (BB/J004561/1).

Contents

List of Tables

List of Figures

Introduction

Secondary metabolites: definition and biosynthetic origins

Secondary metabolites, also referred to as specialised metabolites, are small molecules originating from the metabolism of actinomycetes, non-actinomycete bacteria, filamentous fungi, higher plants and animals that serve survival functions for the organisms producing them (Demain and Fang, 2000). Secondary metabolites are not essential for normal growth, development or reproduction of that organism and have restricted distribution, being distinctive of a genus, species or even a strain (Vaishnav and Demain, 2011; McMurry, 2011; Herbert, 1989).

According to the biochemical origin of secondary metabolites, they are classified into five categories: terpenoids and steroids, fatty acid-derived and polyketides, alkaloids, nonribosomal peptides and enzyme cofactors (McMurry, 2011). In the case of bacteria, the biosynthesis of secondary metabolites is observed during the late-growth phase, also known as idiophase, through specialised biochemical pathways that use biosynthetic units derived from primary metabolism, synthesised during the tropophase. The most usual biosynthetic units used in the biosynthesis of secondary metabolites are acetyl-CoA, shikimic acid, mevalonic acid, methylerythritol phosphate or amino acids (Dewick, 2009).

The specialised biochemical pathways devoted to the biosynthesis of secondary metabolites are encoded in the genome or plasmids of the producer organism. The genes of a specialised biochemical pathway are usually clustered together as a gene cluster. These clusters have been defined as self-contained cassettes for secondary metabolite production (Osbourn, 2010) and consist of groups of physically linked genes that are functionally related and co-expressed and, unlike operons, the genes within these clusters are transcribed separately (Kliebenstein and Osbourn, 2012; Nett, 2014). The genes located within a biosynthesis gene cluster encode for the enzymes that perform numerous tasks that lead to the final metabolic product. Examples of these tasks are biochemical transformations of biosynthetic units, transcriptional regulation of key genes, auto-resistance and/or transport/export of the metabolite produced outside of the cell, and incorporation of post-synthetic modifications to the precursor molecule, that give rise to the final sec-

ondary metabolite (Bibb, 2005).

Surprisingly, the genome of a microorganism may encode for a great number of biosynthesis gene clusters whose potential product might not be produced or detected under standard laboratory conditions (Baltz, 2008). The expression of a biosynthesis gene cluster in the cell is produced under certain environmental conditions such as, depletion of carbon, phosphate or nitrogen sources that, as a consequence, trigger metabolic shifts in the cell that lead to the production of secondary metabolites for self-defence purposes, for example (Martin and Liras, 1989).

Notwithstanding the limited number of biosynthetic units available to assemble secondary metabolites, the variety of chemical structures known among organisms is vast. This is directly associated with the multiplicity of arrangements and variations of the genes present within a biosynthesis gene cluster that link together biosynthetic units in different combinations to give rise to different classes of molecules. Conveniently for us, these molecules present useful applications.

Application of secondary metabolites

Secondary metabolites play an important role in human society due to the wide range of biological activities that they display. They can be used as immunosuppressant, antihelminthic, anticancer, antifungal and antimicrobial, among other advantageous applications for agriculture, food and medical industries. One group of secondary metabolites that has contributed to improving human health are antibiotics. Their discovery at the beginning of the twentieth century meant the cure of life-threatening diseases produced by pathogenic bacteria (Vaishnav and Demain, 2011).

Antibiotic molecules are produced by all kinds of organisms (Figure 1); they represent two-thirds of known secondary metabolites. By far, actinomycetes are the largest producers of antibiotic molecules (7,900), followed by higher plants (5,000) and fungi (2,600)

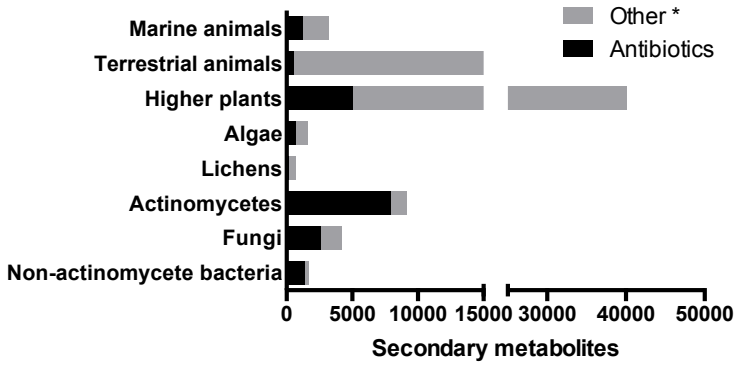


Figure 1: Number of secondary metabolites produced by type of organism. *Secondary metabolite with biological activity different from antimicrobial.

(Kieser et al., 2000). The soil-dwelling *Streptomyces* genus belongs to the actinomycete family and is responsible for producing many of the clinically relevant antibiotics, of which chloramphenicol (*Streptomyces venezuelae*), oleandomycin (*Streptomyces antibioticus*) and vancomycin (*S. orientalis*) are three examples. Of all antibiotics produced by actinomycetes, 75% is synthesised by *Streptomyces*. During the golden era of antibiotics (1950 –1960) nearly 80% of antibiotic compounds were isolated from *Streptomyces*, making this a very prolific genus (Bérdy, 2005).

Antibiotics: definition, applications and antimicrobial resistance

Antibiotics are naturally occurring secondary metabolites that are synthesised by microorganisms, such as actinomycetes, which are capable of inhibiting growth or killing microorganisms at very low concentration (Yoneyama and Katsumata, 2006). The general mode of action of antibiotics is through specific interaction with a bacterial target(s), which are different from those present in eukaryotic cells; therefore, when they are used as drugs, antibiotics do not display (significant) toxicity to the human or animal host.

The introduction of antibiotics has been one of the most revolutionary advancements in human medicine; low concentration of an antibiotic is enough to selectively treat an infectious disease, killing susceptible pathogens and saving lives. However, their misuse and overuse has led to antibiotic-resistant bacteria, making drugs less effective or even ineffective to treat the same infection, forcing the use of stronger and more toxic drugs.

Antimicrobial resistance is defined by the World Health Organization (World Health Organization, 2014) as the resistance of a microorganism to an antimicrobial drug that was originally effective for treatment of infections caused by it. The reasons for the emergence of antibiotic-resistant bacteria are associated with random point mutations in key genes, horizontal transfer of resistance genes through plasmids; the intrinsic resistance of a pathogen, due to the presence of resistance genes in the bacterial genome or increased copy number of the target gene, making the intracellular concentration of the antibiotic insufficient to produce the antimicrobial activity (Davies and Davies, 2010). In addition, multidrug efflux pump systems of the major facilitator superfamily (MFS) represent common mechanisms for bacterial resistance to antimicrobial agents (Kumar et al., 2013).

The need for new antibiotics

In terms of healthcare, some infections triggered by antibiotic-resistant bacteria may be life threatening or even lethal, because of the lack of available treatment for those infections. In addition, the number of approvals of new antibiotics has been progressively declining in the last decades in the United States (Centers for Disease Control and Prevention,

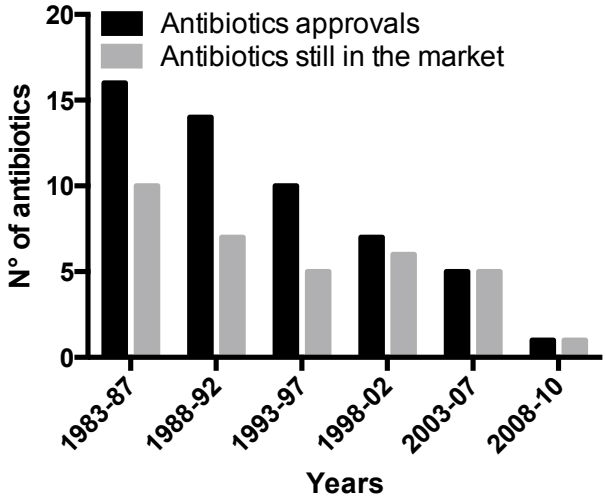


Figure 2: Number of antibiotics approvals by the FDA to treat systemic infections. Adapted from (Outterson et al., 2013; Outterson, 2014).

2013). This scenario is more complex if we consider the fact that some of the approved drugs have been subsequently withdrawn or discontinued from the market, either for safety reasons or they were neither commercially nor clinically successful (Outterson et al., 2013) (Figure 2). Other antibiotics are no longer effective, due to the emergence of resistance that leaves a few or, in the worst case, no treatment options for some infections, such as those produced by VRE (Energy and Commerce Committee, Health Subcommittee United States House of Representatives, 2014).

In the United States, at least 2 million people acquire severe infections triggered by antibiotic-resistant bacteria; of those, at least 23,000 die each year. In 2011, MRSA caused the death of 11,285 people, while VRE caused the death of 1,300 people (Centers for Disease Control and Prevention, 2013). According to the Public Health Institute (*Instituto de Salud Pública*) of Chile, between 2010 and 2012 the average number of MRSA strains isolated from hospitals reached 100 per year, while 779 of VRE were isolated per year (Figure 3; Instituto de Salud Pública de Chile, 2013b; Instituto de Salud Pública de Chile, 2013a).

Today, antibiotic-resistance has been declared a serious and real menace in public healthcare worldwide, which urgently needs the development of new antibiotic drugs to fight against the unavoidable emergence of antibiotic-resistant bacteria (Energy and Commerce Committee, Health Subcommittee United States House of Representatives, 2014).

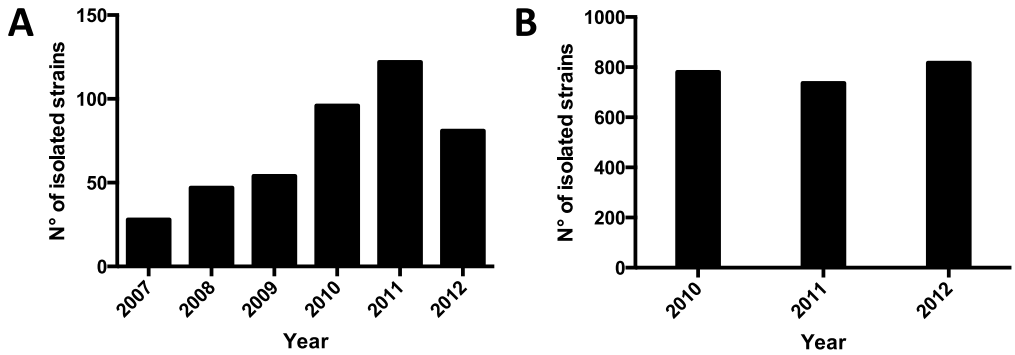


Figure 3: Number of antibiotic-resistant strains isolated from Chilean hospitals per year. (**A**) Methicillin-resistant *Staphylococcus aureus*-community associated; (**B**) vancomycin-resistant *Enterococcus*.

Where to look for new drugs to fight against antibiotic-resistant bacteria?

Historically, secondary metabolites synthesised by microorganisms isolated from natural environments have been the primary source of clinically relevant molecules used in antibiotic development.

However, in the last decades, pharmaceutical companies have abandoned this pipeline for technical (continuous re-discovery of known molecules) and economic reasons (high costs of screening). Instead, they have created large synthetic chemical libraries to screen for candidates molecules or leads (Bérdy, 2005), but they have not been very successful in the field of antibiotics (Payne et al., 2007; Bologa et al., 2013). A shortcoming of these libraries is that they capture only a tiny fraction of the chemical diversity space, conversely, libraries inspired by secondary metabolites bring more complex and biologically optimised molecules, which may increase the chances of finding novel leads (Shoichet, 2013).

To find novel drug candidates, both marine and soil environments have been recently explored. In particular, bioprospecting in extreme environments like the Atacama Desert in northern Chile surprisingly resulted in finding great microbial diversity, such as members of the *Amycolatopsis*, *Lechevalieria* and *Streptomyces* genera (Okoro et al., 2009; Bull and Asenjo, 2013). Great metabolic diversity related to those bacteria found has also been reported recently (Rateb et al., 2011a,b; Nachtigall et al., 2011).

Nowadays, there is a renaissance in exploring, or bioprospecting unexplored environments to find new microorganisms and novel chemical scaffolds. Improvements in technologies such as genome sequencing, bioinformatics and synthetic biology, have allowed scientists to unveil novel biochemical pathways from new microorganisms more efficiently (Bologa et al., 2013).

General characteristics of the *Streptomyces* genus and *Streptomyces leeuwenhoekii* C34 species

Streptomyces is a soil-dwelling Gram-positive, aerophile bacterium and the largest genus belonging to actinobacteria. Usually, *Streptomyces* have a large genome size (about 8 Mbp) with high G+C content. One characteristic feature of this genus is the formation of aerial mycelium when cultivated in the appropriate medium. Moreover, *Streptomyces* is successful in colonising a wide variety of environments since it can grow as hyphae and differentiate to spores; this characteristic facilitates its survival in soil, and other environments, for prolonged periods of time (Kieser et al., 2000).

Streptomyces leeuwenhoekii C34 is a novel species isolated from the Chaxa lagoon from a hyper-arid and salt-flat region with no vegetation, located in "Los Flamencos National Reserve", at 2,300 metres above sea level in the Salar de Atacama, near Tocanao in northern Chile (Okoro et al., 2009). *S. leeuwenhoekii* C34 is capable of producing diffusible pigments, some known molecules such as hygromycin A and desferrioxamine E and two new antibiotics, the chaxamycins and chaxalactins. It forms branched substrate mycelium that carry aerial hyphae, which differentiate into spiral chains of smooth-surfaced spores. It grows from 4 to 50 ℃ and optimally at 30 °C; from pH 6.0 to 11 and optimally at 7.0; it is also able to grow in the presence of 10% w/v sodium chloride. The genome of *S. leeuwenhoekii* C34 has been recently sequenced using the Illumina-sequencing platform, which led to an assembly of 658 contigs for a total genome size of 7.86 Mb, and the G+C content in its genome is 72.6 mol%, consistent with other streptomycetes (Busarakam et al., 2014).

Chaxamycins and chaxalactins as potential antibiotics

Chaxamycins and chaxalactins (Figure 4) are two new groups of polyketides produced by *S. leeuwenhoekii* C34. Both display antimicrobial activity. Chaxamycins belong to the group of ansamycin-type polyketides, where *ansa* in the word ansamycin comes from Latin that means *handle*, and refers to the characteristic cyclic structure comprised of an aromatic moiety and an aliphatic chain that forms a ring through an internal amide bond (Floss et al., 2011; Herbert, 1989). Chaxamycin A displayed anticancer activity against the human Hsp90α enzyme, involved in cancer proliferation, while chaxamycin D has been proven to effectively inhibit growth of MRSA strains; some of those strains were not inhibited by rifampicin, an antibiotic drug derived from rifamycin SV and produced by *Amycolatopsis mediterranei*, used to treat infections produced by MRSA strains (Rateb et al., 2011a). MIC is the minimal concentration of a compound in µg/ml needed to inhibit microbial growth by 50% (MIC_{50}). A promising MIC_{50} value for a compound ranges between <1 –10 µg/ml (Mostafa Rateb, personal communication). Chaxamycin D for instance, showed MIC_{50} values between 0.05-1.20 µg/ml against Gram-negative and MRSA strains (Rateb et al., 2011a).

Chaxalactins A−C, display selective antimicrobial activity against Gram-positive organisms such as *Staphylococcus aureus*, *Listeria monocytogenes* and *Bacillus subtilis*. Compounds similar to chaxalactins are rarely found in nature (Rateb et al., 2011a), which makes them suitable candidates for further studies to improving their antimicrobial activity. As polyketides, both chaxamycins and chaxalactins are synthesised by a type I PKS, but according to their chemical structures they are probably originated from different biosynthesis pathways encoded in the genome of *S. leeuwenhoekii* C34 (Rateb et al., 2011b).

PKS and Polyketides

One important group of secondary metabolites are the polyketides. Polyketides are very important molecules in the drug industry since about 0.3% of the known polyketides have been commercialised, which is 300 times better than the typical <0.001% from pharmaceutical screens (Weissman and Leadlay, 2005). A polyketide molecule is a poly-β-keto chain that is synthesised by a PKS enzyme. Polyketides are incredibly diverse molecules, which strongly depends on the architecture of the PKS that varies from one microorganism to another. In addition, post-synthetic modifications made on the polyketide backbone increase this diversity. The biosynthesis of polyketides resembles that of fatty acids. A PKS catalyses a series of Claisen condensation that assembles units of acetyl-CoA with other acyl-CoA-derivatives such as malonyl-CoA, propionyl-CoA, ethylmalonyl-CoA and others to give rise to a polyketide chain that is usually cycled to yield a macrolactone ring (Dewick, 2009).

There are several types of PKSs. Type I are known as non-iterative multi-modular enzymatic complexes that catalyse the elongation of a polyketide chain by adding extender units in one cycle; type II are separable units that carry out the elongation of the polyketide chain in an iterative fashion (Shen, 2003); type III, also known as chalcone synthase, are more simple homodimeric enzymes (a homodimer of identical KS domains) that are iteratively acting condensing enzymes which use a larger set of starter units in comparison to type I and II PKSs (Shen, 2003; Austin and Noel, 2003).

A PKS is organised in modules, each module harbours several domains that perform a specific task as shown in Figure 5 (Keatinge-Clay, 2012). AT is a monomeric domain of 300 amino acids that includes subdomains α/β-hydrolase fold (240 amino acids) and

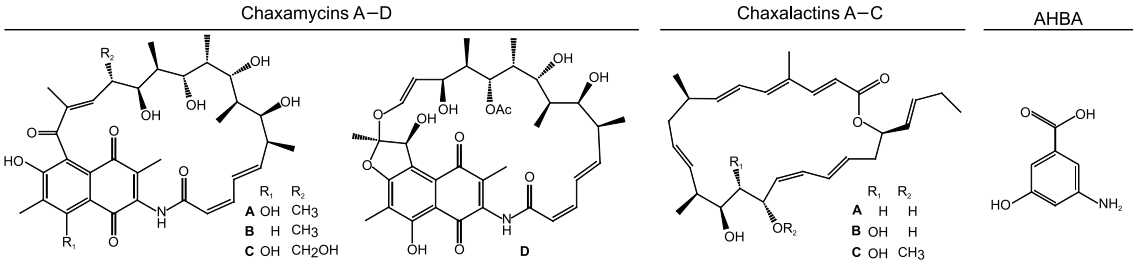


Figure 4: Structures of polyketides chaxamycins A−D and chaxalactins A−C, and AHBA.

ferredoxin-like (60 amino acids), they are involved in the selection of acyl-CoA extender units. The specificity of an AT domain could be predicted by assessing the presence of amino acid motifs: if YASH motif is present, then AT will recognise a (2*S*)-methylmalonyl-CoA; if motif HAFH is present, then AT will recognise malonyl-CoA as the extensor unit. The presence of the catalytic serine residue in GH**S**xG motif indicates that AT is active (Smith and Tsai, 2007; Dunn et al., 2013). ACP is an 80-residue domain located at the C-terminus of each module, it comprises four-helix structures that harbour a 4'-phosphopantetheine prosthetic group in a serine residue located in the (D/E)xGxD**S**L motif, which shuttles extender units and polyketide intermediates from one module to another. KS is a dimeric domain with a thiolase fold that catalyses the formation of carbon-carbon bonds between an extender unit and a polyketide intermediate through Claisen condensation. Its catalytic triad is comprised of a cysteine and two histidine residues present in TA**C**SSS, EA**H**GTG and KSNIG**H**T motifs, respectively. These domains are essential for the assembly of the backbone of the polyketide chain.

Other domains present in a module of a PKS introduce modifications to the growing polyketide chain. KR is a 460 amino acid monomeric domain containing a Rossmann-like fold that reduces the β-keto group in polyketide intermediates, using the reducing power from NADPH, to give rise to a β-hydroxyl group. KR domains control the stereochemistry of both the α-substituent and the resulting β-hydroxyl group of a polyketide intermediate. Active KR domains can be classified into A−type, if generate a β-hydroxyl in the L-orientation or B−type, if generate a β-hydroxyl in the D-orientation. Inactive KR domains, those that don't reduce the β-ketone group, are classified as C−type. DH is a 280 amino acid domain containing a double hotdog fold that catalyses the dehydration of a β-hydroxyl group in the polyketide intermediate, leaving an unsaturation in the α,β position, with loss of the orientation of the α-substituent. DH domains are also involved in the final configuration of the resulting double bond: if a DH domain acts on a product of a B−type KR domain, then the double bond will be in the *cis* (*Z*) configuration, but if this acts on a product of an A−type KR domain, then the double bond will be in the *trans* (*E*) configuration. The catalytic tyrosine and histidine residues are present in the **Y**AAN and **H**xxxGxxxxP motifs of an active DH domain, respectively.

ER is a 310 amino acid domain that is involved in the reduction of the α,β double bond of a polyketide intermediate using NADPH (enoyl-reduction), modifying the stereochemistry of the α-substituent of this intermediate. After the enoyl-reduction reaction occurs, the configuration of the α-carbon of the polyketide intermediate is modified to either a (2*S*)-configuration or a (2*R*)-configuration. This could be predicted by the presence or absence of a conserved tyrosine residue in the amino acid sequence of the ER domain: if this tyrosine residue is present, then the ER domain is predicted to leave the α-carbon of the polyketide intermediate in the (2*S*)-configuration. Competent ER domains have a conserved NADPH-binding motif, HxAx(G/T)GV(G/A)(M/S)A, essential for the enoyl-reduction (Keatinge-Clay, 2012; Kwan et al., 2008).

Finally, TE is a 240-290 amino acid domain, located at the C-terminus of the last module of the PKS (also referred to as the *End* module) that catalyses the intramolecular cyclisation of the polyketide chain, bonded to the final module of the PKS, through an

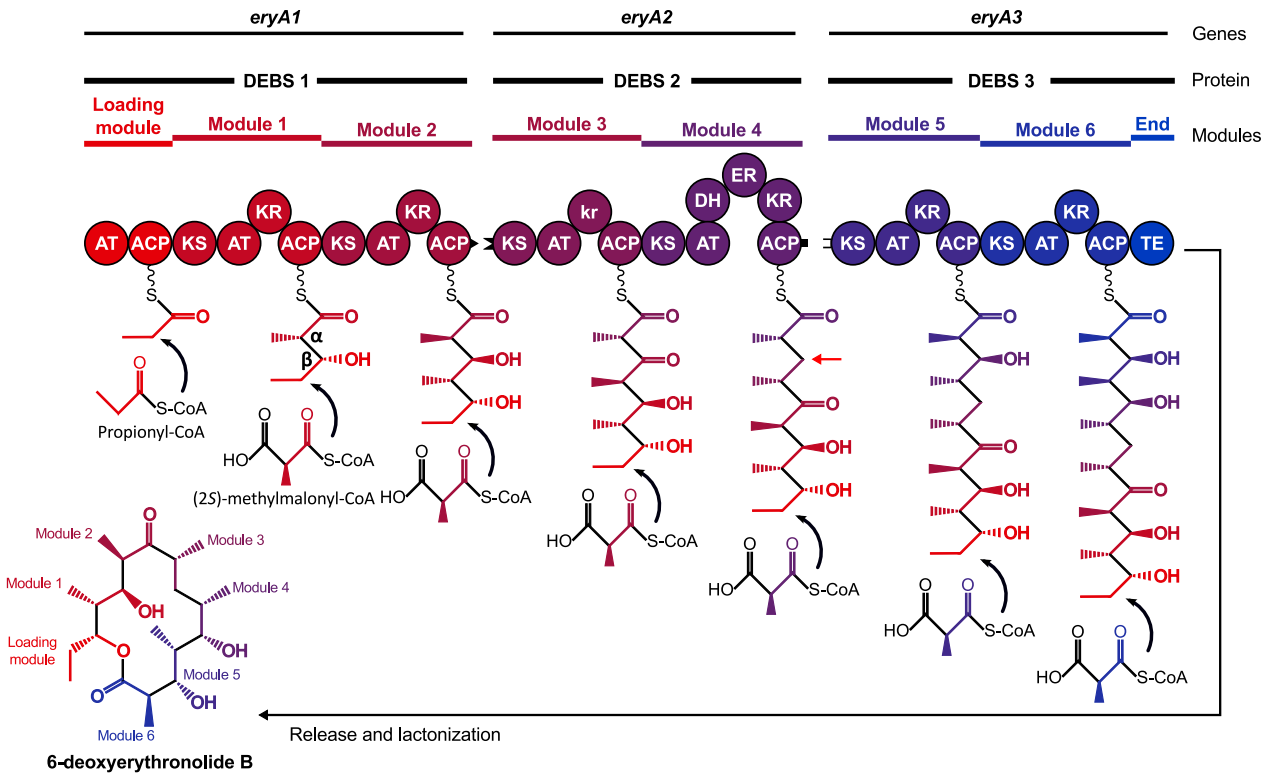


Figure 5: Schematic representation of the mechanism of biosynthesis of the polyketide, 6-deoxyerythronolide by the type I PKS, DEBS. kr in Module 3 is inactive. 4'-phosphopantetheine prosthetic group of ACP is shown as a wavy line. α-substituent and β-hydroxyl groups are indicated with the respective Greek letter; red arrow indicates the position of the β-keto group fully reduced by KR, DH and ER of Module 4. Figure adapted from (Kapur et al., 2012).

intramolecular ester lactone bond, giving rise to a macrolactone ring. A serine residue located in the Gx**S**xG motif is part of the catalytic triad comprised of histidine and aspartate residues (Keatinge-Clay, 2012).

One example of type I PKS is DEBS from *Saccharopolyspora erythrea*, which produces 6-deoxyerythronolide B, the precursor of the erythromycin antibiotic (Figure 5). DEBS is comprised of three protein subunits: DEBS 1, DEBS 2 and DEBS 3, that are encoded in the genome of *Sacc. erythrea* in genes: *eryA1*, *eryA2* and *eryA3* (Dewick, 2009). DEBS 1–3 create a linear sequence of six modules that contain several domains. Each domain catalyses a specific function (explained above) while processing the growing polyketide chain (Bevitt et al., 1992).

The biosynthesis of 6-deoxyerythronolide B (Figure 5) begins in the Loading Module of DEBS with the load of the propionyl-CoA starting molecule into the ACP domain. The Loading Module is comprised of domains AT and ACP, but is devoid of a KS domain, which indicates that only non-carboxylated acyl-CoA species, like propionyl-CoA, and not methylmalonyl-CoA will be the substrate of this Module (Bisang et al., 1999; Wiesmann et al., 1995; Pieper et al., 1995). In Module 1, KS catalyses Claisen condensation that links together a unit of (2*S*)-methylmalonyl-CoA, which has been recognised by the AT of Module 1, with the propionate coming from the ACP of the Loading Module, giving rise to

a diketide. Afterwards, the KR domain of Module 1 reduces the β-ketone group of the propionyl moiety of the diketide (Dewick, 2009). This process of extension carries on sequentially until reaching TE domain, which catalyses the releasing of the polyketide chain and promotes the formation of the macrolactone ring that gives rise to 6-deoxyerythronolide B (Keatinge-Clay, 2012). Module 4 of DEBS 2 contains KR, DH and ER domains that sequentially perform the reduction of the β-ketone group of the pentaketide by the KR domain, the dehydration of the resulting β-hydroxyl group by the DH domain, forming a double bond, and the reduction of this double bond by the action the ER domain (Figure 5, arrow).

Mechanism of biosynthesis of polyketides

Extensive studies have been conducted on PKS domains to understand the mechanism of extension and modification of polyketides (Dewick, 2009; Staunton and Weissman, 2001; Keatinge-Clay, 2012, 2007; Floss and Yu, 1999). The knowledge of PKS was used to compose a schematic representation of the typical reactions that take place in a module of a PKS as shown in Figure 6.

First, KS domain located in the *Current Module* of a PKS catalyses the transfer of the polyketide intermediated from the ACP of the *Previous Module* (Figure 6A). The AT domain of the *Current Module* recognises the next extender unit that will be incorporated into the polyketide chain, which in this example is (2*S*)-methylmalonyl-CoA (Figure 6B). In general, AT domains recognise malonyl-CoA, (2*S*)-methylmalonyl-CoA, etc., and evidence demonstrates that the 2*S* configuration of intermediates is preferred over its 2*R* epimer (Wiesmann et al., 1995). The methylmalonyl moiety attached to the AT domain (Figure 6C) is transferred to the ACP domain in the C-terminus of the *Current Module* and remains there until the next reaction (Figure 6D). Afterwards, the KS of the *Current Module* performs Claisen condensation between the polyketide intermediate attached to it and the methylmalonyl extender unit bound to ACP; this reaction releases CO_2 from the methylmalonyl moiety and also inverts the orientation of the α-substituent from L to D (Keatinge-Clay, 2012) (Figure 6E→F). This is one of the most important reactions in polyketide biosynthesis.

The presence of KR, DH and ER domains in a module of a PKS contributes to increasing the diversity of chemical structures of polyketides, by modifying intermediates in each step. KR domain catalyses the reduction of the β-ketone group to hydroxyl using NADPH; this domain also controls the stereochemistry of the α-substituent and the β-hydroxyl (Figure 6F→G, the zigzag lines in the resulting polyketide intermediate indicate that that bond could be in L or D orientation, depending on the type of KR domain that participated in the keto-reduction reaction). DH domain performs the dehydration of the β-hydroxyl group, leaving a double bond in the α,β-position (Figure 6G→H). There are some DH domains that perform the shift or isomerisation of the double bond from the α,β-position to the atypical β,γ-position, which also contributes to increasing the diversity of chemical structures of polyketides (not shown). ER domain reduces the double bond by using NADPH and

also controls the orientation of the α-substituent, leaving it in either L or D orientation (Figure 6H→I). This process is repeated in the *Next Module* of the type I PKS that might have different domains in it that process the intermediate differently (Figure 6I→J).

When the extension of the polyketide chain reaches the final module, the final polyketide intermediate is shuttled to the TE (Figure 6I→K) that carries out the lactonisation of the polyketide through the secondary alcohol formed at the beginning of the polyketide extension (Figure 6K→L, the hydroxyl group in red represents the secondary alcohol). This finalises with the formation of an intramolecular ester lactone bond in the polyketide intermediate to give rise to a macrolactone (Figure 6L→M).

This schematic representation of the reactions that take place in a PKS will serve as reference for identifying possible products synthesised by a PKS.

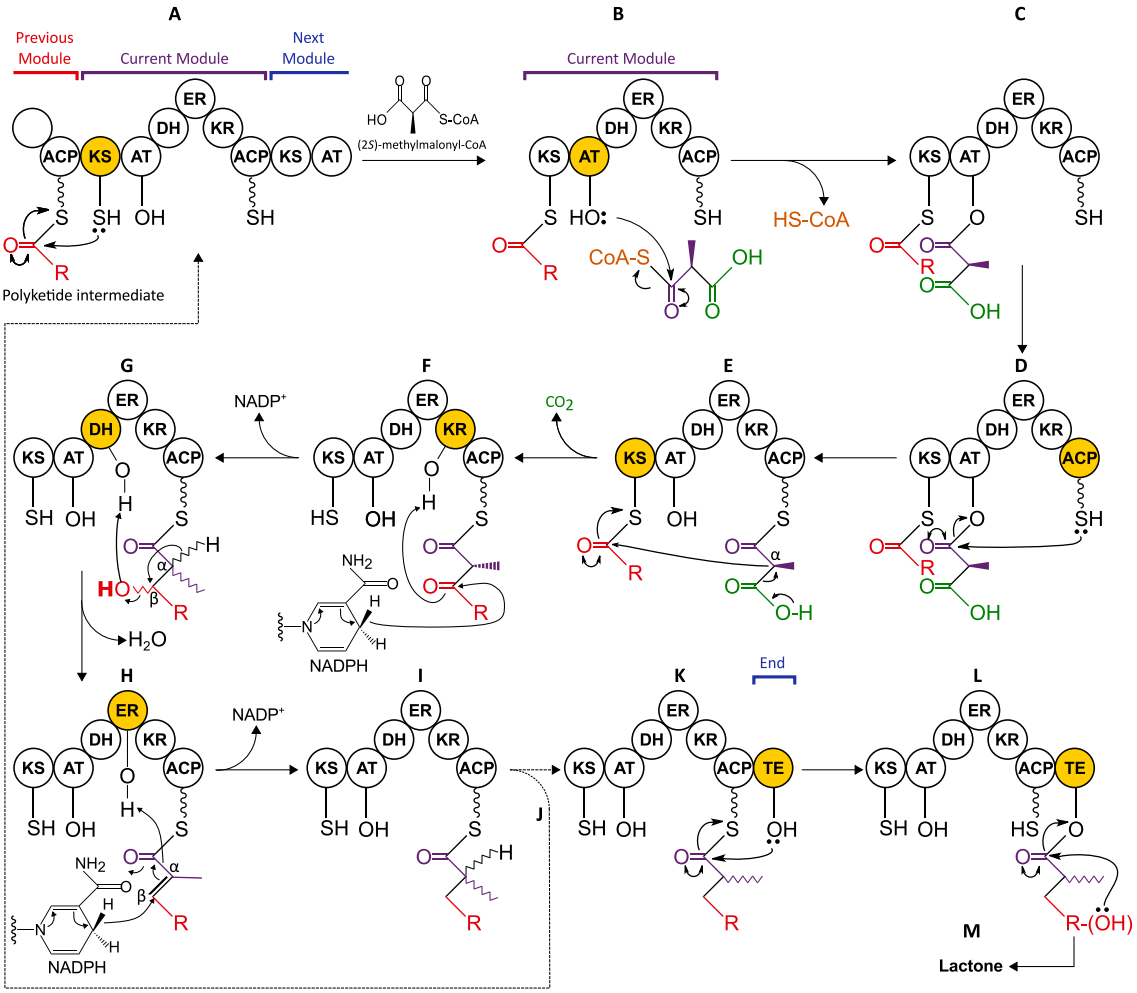


Figure 6: General mechanism of assembly a polyketide by a type I PKS. For the description of each step, please read the main text. The 4'-phosphopantetheine prosthetic group of ACP is shown as a wavy line; zigzag line represents unknown orientation (L or D) of the α-substituent or β-hydroxyl group. Figure adapted from (Dewick, 2009) and complemented with information from (Keatinge-Clay, 2012).

Tools for the identification of biosynthesis gene clusters in bacterial genomes

Bacterial genomes, and especially those of actinobacteria, harbour metabolic pathways devoted to the biosynthesis of secondary metabolites, many of them are likely producers of novel compounds with useful biological activities. Sequencing bacterial genomes has became in a useful strategy towards the identification of secondary metabolite biosynthesis gene clusters. Two examples of sequencing technologies are the well-established Illumina MiSeq (Illumina Inc., 2010) and the new system developed by Pacific Biosciences, PacBio RS II SMRT (Pacific Biosciences, 2015). Bioinformatic tools available on the web allow the annotation of entire genomes in a matter of hours, which give access to have an overview of the function of the genes encoded in a genome. Examples of such tools are the Rapid Annotation using Subsystem Technology, RAST (Aziz et al., 2008), or the Bacterial Annotation System, BASys (Van Domselaar et al., 2005).

The identification of the genes responsible of the biosynthesis of secondary metabolites is a challenging task that usually requires combining bioinformatics, genetic tools and natural product chemistry, strategy known as “Genome mining” (Gomez-Escribano et al., 2015). Bioinformatic tools devoted to the identification of putative biosynthesis gene clusters are the antibiotic and Secondary Metabolite Analysis SHell, antiSMASH 3.0 (Weber et al., 2015), that allows the identification of known biosynthesis gene clusters and the detection of putative gene clusters of unknown types, and the PRediction Informatics for Secondary Metabolomes, PRISM (Suzek et al., 2007), that allows the prediction of genetically encoded nonribosomal peptides and type I and II polyketides. In addition, the search for biosynthesis gene clusters is aided by analysing protein sequences against the NCBI CDD (Marchler-Bauer et al., 2015), UniProt Reference Clusters, UniRef (Suzek et al., 2007), PROSITE scan for functional protein sites (Sigrist et al., 2012), among other tools.

After identifying a putative biosynthesis gene cluster, genetic tools come into action to validate the hypothesis generated from *in silico* analyses. Cloning of entire gene clusters can be carried out in special vectors like the *E. coli–Streptomyces* artificial chromosome, ESAC (Sosio et al., 2000), that allows cloning large fragments of genomic DNA and its transfer from *E. coli* into *Streptomyces* strains. Heterologous hosts derived from *S. coelicolor* have been developed by means of deleting its endogenous biosynthesis gene clusters, increasing the precursors availability, and allowing the identification of novel products of biosynthesis gene clusters in a clean metabolic background. Gomez-Escribano and Bibb (2011) have developed a strain derived from *S. coelicolor* M145, called M1152 and M1154, whose utility is for the heterologous expression of genes cluster from other *Streptomyces* strains. This has been successfully tested by cloning the chloramphenicol and congocidine biosynthesis gene clusters in either *S. coelicolor* M1152 or M1154, resulting in the production of the respective metabolic product.

Genome mining analysis of *Streptomyces* strains has resulted in the discovery of a plethora of putative secondary metabolite biosynthesis gene clusters. For example, the

sequencing of the genome of *S. coelicolor* A3(2) was a major breakthrough in the area of natural product discovery. The biosynthesis gene clusters responsible of producing actinorhodin, prodiginine, calcium-dependent antibiotic and grey spore pigments were known by the time, but further 18 biosynthesis gene clusters were found encoded in its genome (Bentley et al., 2002). Genome mining analysis of *S. coelicolor* led to the discovery of a cryptic type I PKS, responsible of the production of coelimycin P1 (Pawlik et al., 2007; Gomez-Escribano et al., 2012). Other recent results are the discovery of biosynthesis gene clusters responsible of producing the known hygrocin antitumor antibiotics (Li et al., 2014) and the novel neoansamycin antibiotics by *Streptomyces* sp. LZ35 (Li et al., 2015). Chaxapeptin is a novel lasso-peptide with anticancer activity against human lung cancer cell line A549, produced by *S. leeuwenhoekii* C58, also identified by genome mining analysis (Elsayed et al., 2015). Thus, this strategy is becoming a standard and successful procedure for the discovery of novel natural products from genomes.

Aims and description of thesis

The aims of this thesis are to identify the biosynthesis gene clusters of chaxamycins and chaxalactins, two new antibiotics produced by *S. leeuwenhoekii* C34. The chemical structures of these compounds are known, but their respective biosynthesis routes remain to be elucidated. To achieve this, basic procedures for growth and genetic manipulation of *S. leeuwenhoekii* C34 will be developed and *de novo* sequencing of the genome of *S. leeuwenhoekii* C34 will be conducted by combining two sequencing technologies, Illumina MiSeq and Pacific Biosciences. This will provide a highly accurate genome sequence to perform genome mining studies, with the objective of identifying the chaxamycin and chaxalactin biosynthesis gene clusters. In the case of the chaxamycin biosynthesis gene cluster, experimental verification of the bioinformatic hypothesis will be carried out by genetic modification of *S. leeuwenhoekii* C34 and heterologous expression in *S. coelicolor* M1152. Identification of the gene clusters of chaxamycins and chaxalactins will provide new insights into their biosynthetic pathways that will support future metabolic engineering strategies aiming at increasing the production of these two new antibiotics in the native strain or in a heterologous host.

Objectives

General objective

To identify the biosynthesis routes of chaxamycins and chaxalactins of *S. leeuwenhoekii* C34.

Specific objectives

1. To identify the chaxamycin and chaxalactin biosynthesis gene clusters by bioinformatic means.

2. To establish procedures for the transfer of foreign DNA from *E. coli* ET12567/pUZ8002 to *S. leeuwenhoekii* C34.

3. To identify the chaxamycin biosynthesis gene cluster in the native strain by the deletion of the AHBA synthase gene (*cxmK*).

4. To produce chaxamycins A−D in a heterologous host.

Chapter 1

Establishment of practical procedures for growth and genetic manipulation of *Streptomyces leeuwenhoekii* C34

1.1 Abstract

Chapter One details the development of basic procedures for growth and genetic manipulation of *S. leeuwenhoekii* C34. Firstly, the temperature for the highest chaxamycin production in liquid culture was 30 °C and the temperature for the highest sporulation level on solid medium was 37 °C. Secondly, methods for genetic modification of this strain, which did not exist, had to be developed for the study of biosynthesis gene clusters. *S. leeuwenhoekii* C34 was sensitive to antibiotics used to select genetically modified bacteria: thiostrepton, apramycin, hygromycin B and kanamycin, while it was resistant to nalidixic acid. Genetic modifications of *S. leeuwenhoekii* C34 were successfully carried out by intergenic conjugation with *E. coli* ET12567/pUZ8002 carrying pSET152. Optimisation of the conjugation process indicated that supplementation with either 120 mM $MgCl_2$ or 60 mM $CaCl_2$ significantly increased the conjugation frequency (number of *S. leeuwenhoekii* C34 exconjugants per number of donor *E. coli* cells), compared to that with no salt addition. A molecular toolbox for *S. leeuwenhoekii* C34 was generated. The high-copy temperature-sensitive vector pGM1190, that carries the apramycin resistance gene, was transferred into *S. leeuwenhoekii* C34, giving rise to strain M1661: at 30 °C, M1661 was able to grow in the presence of apramycin, while at 34 °C, 37 °C or 43 °C, it was not able to grow in the presence of apramycin, demonstrating that pGM1190 was eliminated at these temperatures, reverting to the wild-type phenotype that is apramycin sensitive. Although some persistence of the vector was observed at 34 °C, 37 °C was chosen as the temperature to ensure the elimination of pGM1190. Bacterial attachment sites recognised by the recombination system of bacteriophages ΦC31 and ΦBT1, which lay in cloning vectors pJI12738 and pIJ10257, were found in the genome of *S. leeuwenhoekii* C34. The functionality of the constitutive promoter, *ermE**, was proven in *S. leeuwenhoekii* C34.

1.2 Introduction

Streptomyces leeuwenhoekii C34 is the producer of the antimicrobials chaxamycins and chaxalactins, whose biosynthesis route has become of increasing interest for the necessity of developing novel drugs (Rateb et al., 2011a,b; Bull and Asenjo, 2013). The study of the biosynthesis of secondary metabolites in newly isolated microorganisms, like in *S. leeuwenhoekii* C34, is often hampered by unknown bacterial genetics or unfamiliarity with culture conditions (Kieser et al., 2000; Olano et al., 2008). In the case of actinomycetes, there exist an important number of standard procedures and molecular tools available that would allow the development of specific procedures for growth and genetic manipulation *S. leeuwenhoekii* C34 (Kieser et al., 2000; Shirling and Gottlieb, 1966).

The molecular toolbox of *Streptomyces* species spans a wide range of applications (Kieser et al., 2000). One example is the sequence-specific serine recombination system from the bacteriophages of *S. coelicolor*, ΦC31 (Rausch and Lehmann, 1991) and ΦBT1 (Gregory et al., 2003), which have been extensively used in integrative vectors. These vectors carry the gene of the phage integrase and the sequence of the *attP* site; the integrase catalyses the specific integration of the vector into a homologue sequence to the *attP*, called *attB* site, located in the bacterial chromosome. Vectors that carry this system have also been modified to allow the cloning of a gene of interest whose transcription is driven by strong promoters.

Transcription of genes in *Streptomyces* species may be carried out using strong or inducible promoters. One example of a strong promoter is the constitutive *ermE** promoter, an up-regulated variant of the *ermE* promoter that, in *Saccharopolyspora erythrea*, controls the expression of a 16S rRNA methyltransferase (*ermE*), which confers resistance to erythromycin (Schmitt-John and Engels, 1992; Bibb et al., 1985). One example of an inducible promoter is the thiostrepton-inducible *tipA* promoter of *S. lividans*, which normally induces the expression of four genes (Murakami et al., 1989). *tipA* promoter requires the TipA protein, whose expression is induced by thiostrepton. TipA protein is not always present in the genome of actinomycetes, limiting its use (Kieser et al., 2000).

Vectors for gene replacement are also available. The self-replicative temperature-sensitive pGM1190 is a derivative of the pSG5 from *S. ghanaensis* DSM2932, which is not able to replicate in actinomycetes at temperatures higher than 34 °C (Muth et al., 1989). Other vectors carrying reporter genes have also been developed for use in actinomycetes. *gusA* gene encodes for a GUS enzyme that normally hydrolyses β-glucuronides. *gusA* has been used in vectors coupled to chromogenic assays; if *gusA* is expressed in the host, then GUS is able to hydrolyse the X-Gluc analogue to give rise to 5,5'-dibromo-4,4'-dichloro-indigo, which is a blue precipitate that could be readily observed on agar plates.

The goal of this Chapter is to establish basic procedures for growth and genetic modification of *S. leeuwenhoekii* C34 to study the biosynthesis pathways of chaxamycins and chaxalactins and for future metabolic engineering applications (deletion of genes or overexpression of genes).

1.3 Materials and methods

1.3.1 Strains, plasmids and primers

Lists of strains, plasmids and primers used in this work can be found in Appendix A.1, Appendix B.1 and Appendix C.1, respectively.

1.3.2 Solutions and media recipes

Recipes for preparation of agar media, liquid media, antibiotic and other solutions for microbiology, and buffer solutions can be found in Appendix E.1, Appendix E.2, Appendix E.3 and Appendix E.4, respectively.

1.3.3 Culture conditions for *S. leeuwenhoekii* C34

S. leeuwenhoekii C34 (strain C34^T = DSM 42122T = NRRL B-24963T) was provided on ISP2 agar plate from a culture collection of Prof. Michael Goodfellow (University of Newcastle, Newcastle upon Tyne, UK) (Busarakam et al., 2014). Routine sub-culture of the strain was done in SFM agar medium at 30 °C for 5 days. Spore stocks of *S. leeuwenhoekii* C34 were prepared following standard procedures (Appendix F.3.1).

Growth and sporulation of *S. leeuwenhoekii* C34 was tested in three different media: SFM, DNA and R2. Both SFM and DNA media were supplemented with trace elements at the same concentration as used in R2 medium (Kieser et al., 2000). SFM agar medium without trace elements was used to evaluate sporulation levels of *S. leeuwenhoekii* C34 at three temperatures: 30 °C, 37 °C and 43 °C. Each plate was inoculated with $1{\cdot}10^8$ spores.

Liquid cultures of *S. leeuwenhoekii* C34 were carried out in 250 ml Erlenmeyer flasks, using 50 ml as working volume of modified ISP2 liquid medium. A seed culture of *S. leeuwenhoekii* C34 was used to inoculate fresh medium with 0.2 of initial OD measured at 600 nm (procedure in Appendix F.3.2). Cultures were stirred at 250 rpm and incubated at the desired temperature for 72 h.

1.3.4 Establishment of chaxamycin production conditions in liquid culture for *S. leeuwenhoekii* C34

Chaxamycin production was evaluated in modified ISP2 liquid medium at three temperatures: 30 °C, 37 °C and 43 °C. Samples of 1 ml were withdrawn from the flask at 10 h, 24 h, 48 h and 72 h and immediately centrifuged at 14,000 rpm at 4 °C for 10 min. The measurement of chaxamycins was carried out from the supernatant fraction, since higher

amounts of those compounds were detected in the supernatant (data not shown). The protocol for measurement of chaxamycins is detailed in Appendix F.5.

1.3.5 Test of antibiotic susceptibility of *S. leeuwenhoekii* C34

The susceptibility of *S. leeuwenhoekii* C34 to kanamycin, apramycin, hygromycin B, thiostrepton and nalidixic acid was tested to determine its sensitivity to antibiotics commonly used to select for genetically modified *Streptomyces*. Concentrations from 0.5 µg/ml to 5.0 µg/ml, with increments of 0.5 µg/ml were tested for kanamycin, apramycin, hygromycin B and thiostrepton; concentrations of 2 µg/ml, 5 µg/ml, 10 µg/ml, 15 µg/ml, 20 µg/ml, 25 µg/ml, 30 µg/ml, 40 µg/ml, 50 µg/ml were tested for nalidixic acid. Two ml of SFM agar were placed in 2 ml Eppendorf tube and mixed with a volume of each antibiotic solution to reach the desired concentration; each tube was kept in a heating block set at 50 °C, to avoid solidification of the agar and promote the mixture of the medium and the antibiotic. The solution SFM-antibiotic was transferred into a square Petri dish (10 cm side, with 25 square wells) and each well was inoculated with $1 \cdot 10^8$ spores of *S. leeuwenhoekii* C34 and cultivated at 30 °C for 7 days.

1.3.6 Conjugation between *E. coli* and *S. leeuwenhoekii* C34

The general procedure of mobilisation of plasmid DNA from *E. coli* strains into *S. leeuwenhoekii* C34 is detailed in Appendix F.3.3.

1.3.6.1 Optimisation of the conjugation between *E. coli* ET12567/pUZ8002, carrying pSET152, and *S. leeuwenhoekii* C34

Optimisation of the conjugation conditions between *E. coli* ET12567/pUZ8002, carrying pSET152, and *S. leeuwenhoekii* C34 was carried out by supplementing the SFM agar medium with 10 mM, 20 mM, 30 mM, 60 mM and 120 mM of either $MgCl_2$ or $CaCl_2$. pSET152 carries the apramycin resistance gene, *acc(3)IV*, and integrates into the chromosome of *S. leeuwenhoekii* C34 in the ΦC31 *attB* site, conferring resistance to apramycin. Before the mobilisation of pSET152 into *S. leeuwenhoekii* C34, it needs to be passed through a methylation-deficient strain such as *E. coli* ET12567, since the majority of streptomycetes have methyl-specific restriction systems (Kieser et al., 2000). The ratio between recipient and donor cells was $1.4 \cdot 10^7$ *S. leeuwenhoekii* C34 spores (recipient) and $1.6 \cdot 10^8$ *E. coli* cells (donor). Transformants were selected by overlaying the surface of SFM plates with a solution of 50 µg/ml apramycin plus 20 µg/ml nalidixic acid as indicated in Appendix F.3.3. This gave rise to *S. leeuwenhoekii* strain M1660.

1.3.7 Usage of the temperature-sensitive pGM1190 in *S. leeuwenhoekii* C34

Conjugation between *E. coli* ET12567/pUZ8002 carrying pGM1190 and *S. leeuwenhoekii* C34 was performed as described in Appendix F.3.3. Transformants of *S. leeuwenhoekii* C34 were selected by overlaying the surface of SFM agar plates with a solution of 50 µg/ml apramycin plus 20 µg/ml nalidixic acid, then plates were incubated at 30 °C, a temperature at which pGM1190 is able to replicate in *Streptomyces* (Muth et al., 1989). The resulting strain, *S. leeuwenhoekii* M1661 was cultivated on four plates with SFM agar without antibiotics. Each plate was incubated in parallel at one temperature: 30 °C, 34 °C, 37 °C or 43 °C for 4 days. Loss of pGM1190 was evaluated by streaking the colonies that grew on SFM without antibiotics from the previous step, onto new SFM agar plates supplemented with 50 µg/ml apramycin. These plates were incubated at 30 °C for 3 days. If pGM1190 is eliminated from *S. leeuwenhoekii* M1661 due to the increase in temperature, then this strain will revert to wild-type phenotype, and would not be able to grow in the presence of apramycin.

1.3.8 Agar-based chromogenic assay with *gusA* as reporter gene

S. leeuwenhoekii C34 was transformed with either pGUS (a promoterless pSET152 derivative with the *gusA* reporter gene; Myronovskyi et al., 2011) or pIJ10740 (a pGUS derivative that contains the *gusA* gene downstream to the constitutive *ermE** promoter; Morgan Feeney, personal communication) to yield strains M1655 and M1656, respectively. The expression of *gusA* was evaluated through agar-based chromogenic assays, as described in Sherwood and Bibb (2013). X-Gluc was dissolved in dimethylformamide at 40 mg/ml. Exconjugants of *S. leeuwenhoekii* C34 were overlaid with X-Gluc solution at a final concentration of 160 µg/ml (considering the volume of the plate as 25 ml). Plates were incubated at 30 °C. *S. leeuwenhoekii* C34 wild-type and M1655 were used as controls.

1.3.9 Bioinformatic analysis

Local BLASTn or BLASTp against a database of genome/genes or proteins of *S. leeuwenhoekii* C34 was performed using NCBI package 2.2.26+ (Camacho et al., 2009) and SequenceServer (Priyam et al., in preparation) on a MacBook Pro OS X version 10.7.5. BLAST searches were performed at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul et al., 1997) and UniProt (http://www.uniprot.org/uniprot; Suzek et al., 2007) servers.

1.3.10 Generation of illustrations

ImageJ (http://imagej.nih.gov/ij) version 1.47 and Inkscape (https://inkscape.org) version 0.91 were used to create and edit drawings and illustrations.

1.4 Results

1.4.1 Establishment of growth and sporulation conditions for *S. leeuwenhoekii* C34

1.4.1.1 Selection of temperature and culture medium for fastest growth and sporulation of *S. leeuwenhoekii* C34

Sporulation is an important biological process that, in terms of laboratory procedures, allows the preservation of *S. leeuwenhoekii* C34 strains for future use. Three different media, widely used for other *Streptomyces* strains, were evaluated to select the most appropriate for routine sub-culture and for spore stocks preparation of *S. leeuwenhoekii* C34: R2, DNA and SFM. The formulation of R2 medium contains trace elements and, in order to have comparable results, trace elements were also supplemented to DNA and SFM. The presence of trace elements in culture media markedly increases sporulation levels in other actinomycetes (Mervyn J. Bibb, personal communication).

S. leeuwenhoekii C34 was able to grow at 30 °C in all media tested, however, different sporulation and confluence levels were observed during the study (Figure 1.1). On the third day, more than 90% confluence was reached on SFM, about 70% on DNA medium and about 15% on R2 medium (estimated values); only in the case of R2, confluence moderately increased towards the seventh day. The overall sporulation level of *S. leeuwenhoekii* C34 on either DNA or R2 was low, compared to sporulation on SFM. Undoubtedly, the highest sporulation level was observed on SFM agar medium, which was correlated with increased dark pigmentation observed on the surface of the colonies, consistent with the description made for this strain (Busarakam et al., 2014); the pigmentation is more

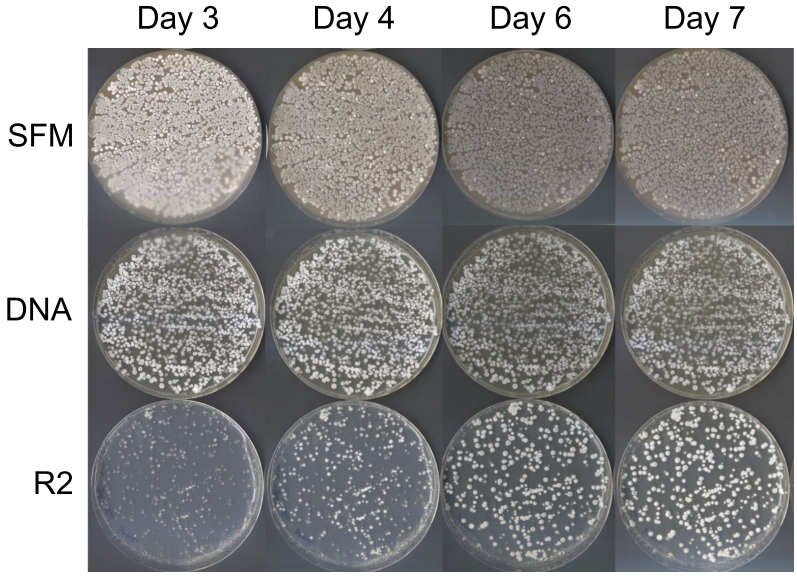


Figure 1.1: Evaluation of growth and sporulation progress of *S. leeuwenhoekii* C34 on different media over time. *S. leeuwenhoekii* C34 was cultivated at 30 °C for 7 days on R2, DNA and SFM media supplemented with trace elements.

noticeable between the fourth and the sixth day. Subsequent tests indicated that it is not necessary to add trace elements to SFM to reach the same sporulation levels with *S. leeuwenhoekii* C34. Thus, SFM allowed the fastest growth and sporulation of *S. leeuwenhoekii* C34 and was selected for routine growth and spore-stock preparation.

Temperature was another parameter assessed to establish growth conditions of *S. leeuwenhoekii* C34. Three temperatures were selected for its cultivation on SFM: 30 °C, recommended as optimal growth temperature (Busarakam et al., 2014), 37 °C and 43 °C, since this bacterium was isolated from the Atacama Desert. The results indicated that *S. leeuwenhoekii* C34 was able to grow and sporulate at all temperatures tested, being more accelerated at higher temperatures (Figure 1.2). Early growth of *S. leeuwenhoekii* C34 was already observed on the first day on plates incubated at 37 °C and 43 °C, being slightly higher at 43 °C. No growth was observed on plates incubated at 30 °C on the first day. Surprisingly, sporulation of *S. leeuwenhoekii* C34 was observed on the second day, on plates incubated at 37 °C and 43 °C, whereas at 30 °C sporulation was observed at the fourth day. In addition, the diameter of the colonies was considerably wider at 43 °C and also, the production of diffusible pigments was more noticeable at 37 °C and 43 °C than at 30 °C.

In summary, temperatures above 37 °C are more suitable to reach copious growth and high sporulation levels in shorter periods of time, which is advisable for spores-stock preparation of *S. leeuwenhoekii* C34.

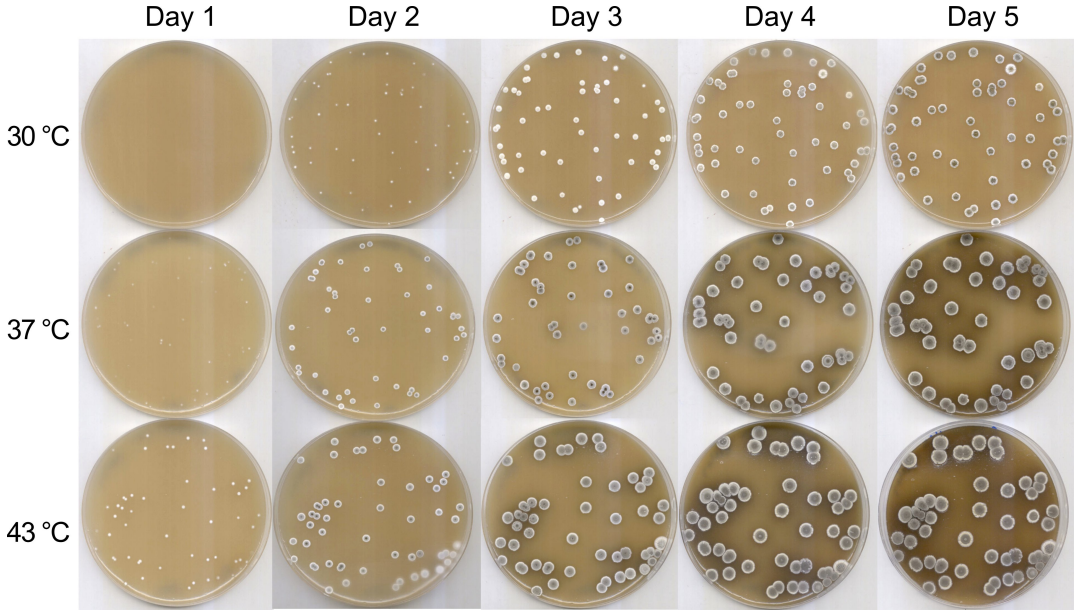


Figure 1.2: Evaluation of growth and sporulation progress of *S. leeuwenhoekii* C34 at different temperatures over time. *S. leeuwenhoekii* C34 was cultivated at 30 °C, 37 °C and 43 °C for 7 days on SFM media without trace elements.

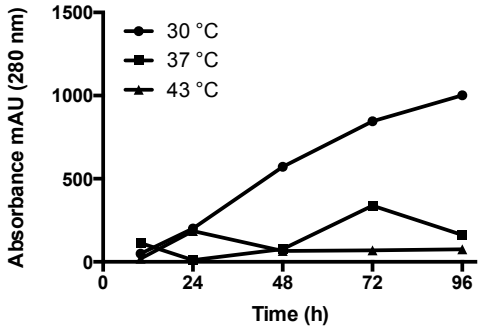


Figure 1.3: Evaluation of chaxamycin production by *S. leeuwenhoekii* C34 at different temperatures.

1.4.2 Establishment of chaxamycin production conditions for *S. leeuwenhoekii* C34

The production of all chaxamycin species (A–D) was evaluated in modified ISP2 liquid medium, at 30 °C, 37 °C and 43 °C to establish a suitable condition to carry out fermentations with *S. leeuwenhoekii* C34. Chaxamycin production was reported as the sum of all species measured in the supernatant fraction at 280 nm. The production level of chaxamycins was considerably higher at 30 °C when compared with that at either 37 °C or 43 °C (Figure 1.3). Thus, all fermentations for evaluating chaxamycin production were performed at 30 °C.

1.4.3 Evaluation of *S. leeuwenhoekii* C34 antibiotic susceptibility

Increasing concentrations of thiostrepton, apramycin, hygromycin B, kanamycin and nalidixic acid were evaluated to test *S. leeuwenhoekii* C34 susceptibility to these antibiotics, commonly used to select for genetically modified actinomycetes (Figure 1.4). Nalidixic acid is used for elimination of *E. coli* after conjugation, therefore it is expected that *S. leeuwenhoekii* C34 is resistant to it. The results indicated that *S. leeuwenhoekii* C34 was not able to grow on SFM supplemented with concentrations >2.5 µg/ml of thiostrepton (Figure 1.4A); >0.5 µg/ml of apramycin (Figure 1.4B); >2.0 µg/ml of hygromycin B (Figure 1.4C); >1.0 µg/ml of kanamycin (Figure 1.4D). Concentrations >25 µg/ml of nalidixic

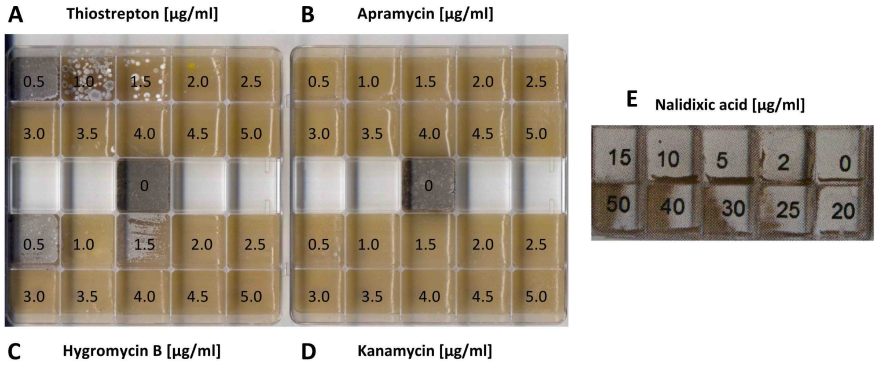


Figure 1.4: Evaluation of *S. leeuwenhoekii* C34 susceptibility to antibiotics commonly used to select for genetically modified actinomycetes.

acid (Figure 1.4E) hampered *S. leeuwenhoekii* C34 growth. Thus, thiostrepton, apramycin, hygromycin B and kanamycin are suitable antibiotics to select for genetically modified *S. leeuwenhoekii* C34 strains.

1.4.4 Conjugation between *E. coli* and *S. leeuwenhoekii* C34

1.4.4.1 Identification of *attB* sites for actinophages ΦC31 and ΦBT1 in the genome of *S. leeuwenhoekii* C34

The sequences of the canonical *attB* sites of ΦC31 (Rausch and Lehmann, 1991) and ΦBT1 (Gregory et al., 2003) described in *S. coelicolor* (Table Appendix G.1) were used to find homologous sites in the genome of *S. leeuwenhoekii* C34 by BLASTn analysis (Camacho et al., 2009). Homologous sequences of *attB* sites were found in the genome of *S. leeuwenhoekii* C34 in homologous genes to those where *attB* sites lay in the genome of *S. coelicolor* (Table1.1; *attB* sequences of *S. leeuwenhoekii* C34 are available in Table Appendix G.1). The homologous *attB* site of ΦC31 was located in the gene *sle_35040* that encodes for a putative pirin-domain-containing protein (Pfam accessions: PF02678 and PF05726) that was 90% identical to *sco3798*. Similarly, the *attB* site of ΦBT1 was located within gene *sle_28260*, which encodes for a putative integral membrane protein that shares 92% of identity with gene *sco4848*.

The presence of both *attB* sites in *S. leeuwenhoekii* C34 suggests that the usage of vectors harbouring integrase systems of ΦC31 and ΦBT1 is feasible.

1.4.4.2 Transfer of pSET152 from *E. coli* ET12567/pUZ8002 to *S. leeuwenhoekii* C34

Conjugation between *E. coli* ET12567/pUZ8002, a methylation-deficient shuttle-vector system, and *S. leeuwenhoekii* C34 was evaluated using vector pSET152 that carries the apramycin resistance gene that integrates into the *attB* site of ΦC31. Preliminary conjugations were conducted following a modified procedure from that reported in Appendix F.3.3: pre-germinated spores of *S. leeuwenhoekii* C34 were allowed to cool for 1.5 h; *E. coli* ET12567/pUZ8002/pSET152 was resuspended in 500 µl of LB and 250 µl of it was mixed with 500 µl of pre-germinated spores; afterwards, increasing volumes were plated on SFM supplemented with 10 mM $MgCl_2$.

Table 1.1: Locus of *attB* sites in genes of *S. leeuwenhoekii* C34 (Sle) and *S. coelicolor* (Sco).

attB site	Locus coordinates	Gene	Predicted gene product function	Accession
ΦC31-Sco	260747..260796	*sco3798*	Pirin protein	1099234
ΦC31-Sle	4154940..4154989	*sle_35040*	Pirin protein	CQR62965[1]
ΦBT1-Sco	201997..201948	*sco4848*	Putative integral membrane protein	1100289
ΦBT1-Sle	3403179..3403228	*sle_28260*	Putative integral membrane protein	CQR62287[1]

[1]Only NCBI protein accession number was available.

As a result, 8 exconjugants of *S. leeuwenhoekii* C34::*pSET152*, named strain M1660, were observed, indicating that pSET152 was apparently integrated into the chromosome of *S. leeuwenhoekii* C34. To verify whether pSET152 was actually integrated into the chromosome, each exconjugant was streaked on SFM supplemented with 50 µg/ml apramycin plus 20 µg/ml nalidixic acid and incubated at 30 °C. Colony-PCR of *S. leeuwenhoekii* M1660 exconjugants was carried out with primers LF044F and LF045R that amplify a region of 118 bp that results after the integration of pSET152 at the ΦC31 *attB* site of *S. coelicolor* (Foulston and Bibb, 2010). The results of the PCR indicate that a band of the expected size was obtained with all exconjugants tested (Figure 1.5, lanes 1−8), whereas no amplification was obtained with genomic DNA of *S. leeuwenhoekii* C34 wild-type (Figure 1.5, lane 9). This result confirms that: mobilisation of plasmid DNA between *E. coli* ET12567/pUZ8002 and *S. leeuwenhoekii* C34 was achieved, *S. leeuwenhoekii* C34 was able to accept foreign DNA, and the integration of pSET152 was located at the expected locus.

1.4.4.3 Effect of $MgCl_2$ and $CaCl_2$ on the conjugation frequency

Initially, conjugations between *S. leeuwenhoekii* C34 and *E. coli* ET12567/pUZ8002 were carried out using SFM agar medium supplemented with 10 mM $MgCl_2$, which is the standard concentration used to improve conjugation performance for streptomycetes (Kieser et al., 2000; Choi et al., 2004). Subsequent experiments suggested that the presence of 10 mM $CaCl_2$ improved the number of transformants obtained after conjugations (data not shown), which has also been pointed out as a parameter that enhances conjugation performance (Wang and Jin, 2014).

Although all conjugations reported in this work were carried out using SFM agar medium supplemented with 10 mM $MgCl_2$ plus 10 mM $CaCl_2$, the individual effect of the presence of both divalent chloride salts, $MgCl_2$ and $CaCl_2$, in conjugations between *E. coli* ET12567/pUZ8002/pSET152 and *S. leeuwenhoekii* C34 was assessed (Figure 1.6). The parameters used in this experiment were: 50 °C for heat-shock of the *S. leeuwenhoekii* C34 spores; the ratio between the recipient and the donor cells was $1.4{\cdot}10^7$ *S. leeuwenhoekii* C34 spores (recipient) and $1.6{\cdot}10^8$ *E. coli* cells (donor). The re-

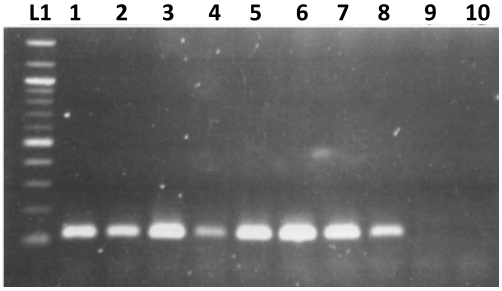


Figure 1.5: Colony-PCR of *S. leeuwenhoekii* M1660 to verify insertion of pSET152 in the genome of *S. leeuwenhoekii* C34. Lane L1 is 100 bp DNA ladder (cat. no. N3231S); lanes 1−8 are *S. leeuwenhoekii* M1660 exconjugants streaked on SFM supplemented with 50 µg/ml apramycin plus 20 µg/ml nalidixic acid; lane 9 is genomic DNA from *S. leeuwenhoekii* C34 wild-type; lane 10 is PCR mixture without template.

sults were expressed as conjugation frequency that was calculated as the number of colonies (exconjugants) obtained on the plate per number of donor cells, which was constant ($1.6 \cdot 10^8$).

The results showed that the addition of either $MgCl_2$ or $CaCl_2$ increased the conjugation frequency, compared to that when no salt was supplemented to SFM. The addition of even 10 mM of either of the divalent chloride salts evaluated increased conjugation performance. The highest conjugation frequency was observed with the addition of 60 mM $CaCl_2$, which was significantly higher compared to that obtained when 60 mM $MgCl_2$ was supplemented. The addition of 120 mM $CaCl_2$ decreased the conjugation frequency to a similar value to that obtained when 30 mM of this salt was supplemented to SFM. This tendency does not occur in the case of addition of $MgCl_2$, since the conjugation frequency increased with increasing concentration of $MgCl_2$, in fact the highest conjugation frequency reached was obtained when supplementing with 120 mM $MgCl_2$. Colony-PCR of exconjugants obtained from plates supplemented with 60 mM $CaCl_2$ showed that all colonies evaluated had the expected amplicon size, indicating that are true exconjugants (Figure Appendix G.1). Based on these results, the supplementation of 60 mM $CaCl_2$ to SFM agar medium provides better conditions for conjugation.

However, the effectiveness of some antibiotics may be altered by the presence of salts. In fact, some antibiotics are more effective at a low concentration of salts, as is the case of kanamycin and hygromycin B (Kieser et al., 2000). For this reason, the antibiotic effectiveness of kanamycin, apramycin, thiostrepton and hygromycin B, in preventing *S. leeuwenhoekii* C34 growth, was evaluated is SFM supplemented with 60 mM $CaCl_2$ (Figure Appendix G.2). The presence of 60 mM $CaCl_2$ did not affect the effectiveness of kanamycin, apramycin and thiostrepton, but it did affect hygromycin B.

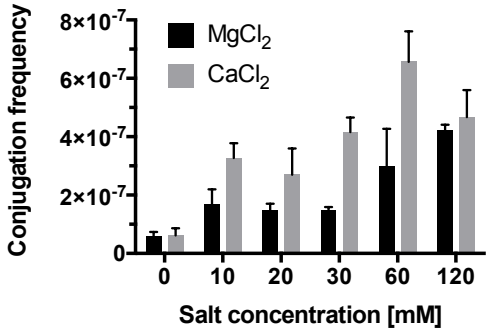


Figure 1.6: Effect of the concentration of $MgCl_2$ and $CaCl_2$ on the conjugation frequency. Values are expressed as conjugation frequency that was calculated as the number of exconjugants obtained per number of donor cells ($1.62 \cdot 10^8$ cells).

1.4.5 Evaluation of the usage of the temperature-sensitive vector, pGM1190, in *S. leeuwenhoekii* C34

pGM1190 is a self-replicative temperature-sensitive vector that cannot replicate in the host at temperatures above 34 °C, it carries the apramycin resistance gene and does not integrate into the bacterial chromosome since it is devoid of an integration system (Muth et al., 1989). This vector has been used before for gene replacement in streptomycetes (Mervyn J. Bibb and Neil Holmes, personal communications), making it a suitable candidate for genetic manipulation of *S. leeuwenhoekii* C34.

Mobilisation of pGM1190 into *S. leeuwenhoekii* C34 was carried out using *E. coli* ET12567/pUZ8002 at 30 °C with subsequent selection of exconjugants with apramycin. The resultant strain that carries pGM1190 was named M1661 (Figure 1.7A). *S. leeuwenhoekii* M1661 was re-streaked on four plates of SFM agar medium without antibiotics; the four plates were incubated in parallel at 30 °C, 34 °C, 37 °C and 43 °C to determine the temperature that eliminates pGM1190 from this strain (Figure 1.7B). For this step, SFM without antibiotics was used because if pGM1190 is eliminated, *S. leeuwenhoekii* M1661 will revert to wild-type phenotype, which is apramycin sensitive. Colonies obtained from plates incubated at different temperatures were sequentially re-streaked on new SFM agar without antibiotics (control) and on new SFM agar supplemented with 50 µg/ml apramycin. All new plates were incubated at 30 °C. If pGM1190 has not been eliminated in the previous step, then pGM1190 can replicate in the host, conferring apramycin resistance to *S. leeuwenhoekii*. The results demonstrated that pGM1190 was eliminated from the host at temperatures above 34 °C, since no growth was observed on SFM agar medium supplemented with apramycin (Figure 1.7C). However, some persistence of pGM1190 was observed when M1661 was incubated at 34 °C in one of the replicates

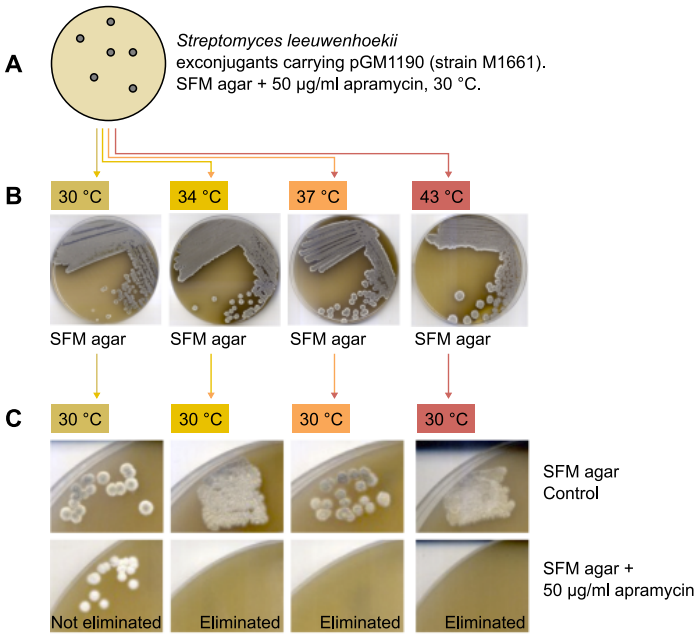


Figure 1.7: Evaluation of the usage of the self-replicative temperature-sensitive vector pGM1190 in *S. leeuwenhoekii* C34. Description in the main text.

(not shown). Therefore, the selected temperature to ensure the elimination of pGM1190 from *S. leeuwenhoekii* C34 strains was 37 °C.

1.4.6 Assessment of the usage of the *ermE** promoter in *S. leeuwenhoekii* C34

Chromogenic assays were used to assess the usage of the constitutive *ermE** promoter in *S. leeuwenhoekii* C34 using *gusA* as reporter gene. In this experiment, the *ermE** promoter drives the transcription of *gusA* that encodes for GUS enzyme. If the *ermE** promoter works in *S. leeuwenhoekii* C34, then *gusA* will be expressed and GUS will degrade X-Gluc, forming a blue precipitate readily observed around the colony (Sherwood and Bibb, 2013). *S. leeuwenhoekii* C34 was transformed with either pGUS or pIJ10740, giving rise to *S. leeuwenhoekii* strains M1655 and M1656, respectively. pGUS is a promoterless vector that carries the *gusA* gene in pSET152 and is normally used as a promoter detection vector (Myronovskyi et al., 2011). pIJ10740 is a derivative of pGUS in which *ermE** promoter drives the transcription of *gusA* (Morgan Feeney, personal communication). Blue halos were only observed around *S. leeuwenhoekii* M1656, which carries pIJ10740 with the *ermE** promoter, and not in either M1655 or *S. leeuwenhoekii* C34 wild-type (not shown). This test validates the usage of the constitutive *ermE** promoter in *S. leeuwenhoekii* C34.

1.5 Discussion

Basic procedures for growth and genetic manipulation of *S. leeuwenhoekii* C34 were established and defined in this Chapter. Growth parameters such as media, temperature and time of incubation were adapted from those reported by Busarakam et al. (2014) for this strain. *S. leeuwenhoekii* C34 was shown to be sensitive to thiostrepton, apramycin, hygromycin B, kanamycin and resistant to nalidixic acid. Procedures for genetic modification that involved the mobilisation of plasmid DNA from *E. coli* ET12567/pUZ8002 to *S. leeuwenhoekii* C34 were successfully established with pSET152, an integrative vector and pGM1190, a self-replicative temperature-sensitive vector. Conjugation performance was significantly improved by supplementing 60 mM $CaCl_2$ to SFM. The usage of the constitutive promoter *ermE** was also confirmed in *S. leeuwenhoekii* C34.

1.5.1 Sporulation and chaxamycin production by *S. leeuwenhoekii* C34 depend upon temperature

S. leeuwenhoekii C34 sporulated on SFM, DNA and R2 media, being more copious and fast on SFM. SFM is described as a medium in which many *Streptomyces* species display effective and good sporulation (Kieser et al., 2000). Temperature also played an important role in the growth of *S. leeuwenhoekii* C34, which was more accelerated at 43 °C than at 30 °C. In general, when a spore of *Streptomyces* encounters the appropriate nutrient and environmental conditions (like temperature) it starts the development of germ tubes that form hyphae. Germ tubes grow and branch to form the vegetative (or substrate) mycelium that creates a net of hyphae that we observe as a colony of *Streptomyces* (Flärdh and Buttner, 2009). This description was readily observed on SFM at 37 °C and 43 °C after a day of cultivation, whereas at 30 °C this was observed on the second day (Figure 1.2).

When the colony depletes nutrients, it starts the production of secondary metabolites and then a second process of differentiation, in which vegetative mycelium grows into the air as aerial hyphae; this structure develops spore walls and synthesises a grey pigment, giving rise to a mature spore (Elliot and Flärdh, 2001; Flärdh and Buttner, 2009). Spores of *S. leeuwenhoekii* C34 were observed on the fourth day at 30 °C, whereas on the second day spores were observed at 37 °C and 43 °C, which supports the idea that temperature accelerates the life cycle of *S. leeuwenhoekii* C34 on solid SFM agar medium. Although the fastest growth was achieved at 43 °C, the agar tended to dry too quickly, making spore stock preparation more difficult, which did not happen at 37 °C. Thus, 37 °C was the preferred temperature to achieve fast sporulation for spore-stock preparation.

On SFM agar medium, grey pigments were observed around colonies (Figure 1.2). This had already been observed for *S. leeuwenhoekii* C34 (Busarakam et al., 2014). Depending on the medium on which *S. leeuwenhoekii* C34 is streaked, different coloured diffusible pigments are described: yellowish diffusible pigments are observed on ISP3, ISP4 and ISP7, grey-yellowish on IPS6 and pale yellow on ISP2, but their metabolic origins remain unknown.

1.5.2 Genetic background for antibiotic susceptibility of *S. leeuwenhoekii* C34

The observed antibiotic susceptibility of *S. leeuwenhoekii* C34 could be explained because it lacks the appropriate resistance determinants encoded in its genome that provide defence in the presence of thiostrepton, hygromycin B, kanamycin or apramycin.

Thiostrepton is a peptide antibiotic produced by *S. azureus* that inhibits the biosynthesis of proteins by binding to the complex 23S rRNA-L11 of the cell. *S. azureus* carries a 2'-*O*-methyltransferase (Tsr; UniProt accession: P18644), which methylates the 23S rRNA, preventing the cell dying by thiostrepton toxicity (Thompson et al., 1982). Four rRNA methyltransferase, homologues to Tsr of *S. azureus*, were found in *S. leeuwenhoekii* C34: Sle_55160, Sle_38790, Sle_55570 and Sle_35420 (NCBI accessions: CQR64973, CQR63338, CQR65014 and CQR63003, respectively). Like Tsr, these proteins contained the SpoU rRNA methylase domain (Pfam accession: PF00588) that is involved in binding the co-substrate, SAM, which carries the methyl group. As expected, none of the rRNA methyltransferases of *S. leeuwenhoekii* C34 harboured the thiostrepton resistance methylase (Pfam accession: PF04705) found in Tsr and needed to confer thiostrepton resistance (Table Appendix G.2). This may explain why *S. leeuwenhoekii* C34 was sensitive to this antibiotic.

S. leeuwenhoekii C34 was sensitive to three aminoglycoside antibiotics, hygromycin B, kanamycin and apramycin. Some microorganisms possess natural mechanisms of defence mediated by enzymes such as those that directly inactivate aminoglycosides such as, the acetyl CoA-dependent AAC, the ATP-dependent ANT, and the ATP-dependent APH (Ramirez and Tolmasky, 2010); or those that modify rRNA targets like site-specific methyltransferases (Wright and Thompson, 1999).

Hygromycin B is produced by *S. hygroscopicus* and inhibits protein synthesis by blocking the translocation step on 70S ribosomes in prokaryotes (and also that of eukaryotes) (Pardo et al., 1985). *S. hygroscopicus* possesses autoimmunity to hygromycin B due to the presence of the gene *hyg* that encodes for the enzyme hygromycin-B 7"-*O*-kinase, Aph(7")-Ia, that harbours an APH family domain involved in the phosphorylation of the 7"-hydroxyl group of the destomic acid moiety of hygromycin B and inactivates it. The genome of *S. leeuwenhoekii* C34 encodes for five APH domain-containing proteins: Sle_01670, Sle_14060, Sle_15890, Sle_24140 and Sle_54290 (Table Appendix G.3), one of them, Sle_01670, shares 87% of identity with Hyg21, a APH enzyme part of the hygromycin A biosynthesis gene cluster of *S. hygroscopicus* NRRL 2388, which has been experimentally verified as a phosphotransferase of hygromycin A (Dhote et al., 2008). A phylogenetic tree of protein sequences was built for a detailed analysis of the relationship of the APHs found in *S. leeuwenhoekii* C34 with those described in the literature (Ramirez and Tolmasky, 2010; Wright and Thompson, 1999) (Figure Appendix G.3). Sle_01670, the homologue to Hyg21, was grouped in the same clade with Aph(4)-Ia and -Ib which, in addition to Aph(7")-Ia, are known to inactivate hygromycin B; Sle_14060 was the closest relative to Aph(7")-Ia, which share 32% of identity. But, it seems unlikely that either Sle_01670 or

Sle_14060 interact with hygromycin B since *S. leeuwenhoekii* C34 was sensitive to it.

Kanamycin is produced by *S. kanamyceticus* and interacts with the 30S subunit of the bacterial ribosome, blocking its function in the cell. *S. kanamyceticus* possesses three known mechanisms to avoid self-intoxication when producing kanamycin, which are encoded in the kanamycin biosynthesis gene cluster: a 16S rRNA methyltransferase, Kmr, that catalyses the methylation of the 16S rRNA to lower the affinity of kanamycin for the antibiotic-binding sites of the ribosome (Demydchuk et al., 1998); a kanamycin-efflux system, Orf7, that mediates the transport of kanamycin outside the cell (Yanai and Murakami, 2004); and an AAC(6'), Kac, that transfers an acetyl group to kanamycin for its inactivation (Flatt and Mahmud, 2007). The genome of *S. leeuwenhoekii* C34 does not seem to encode for a protein similar to Kmr or to Orf7, but in the case of an AAC(6'), it needs a more careful analysis.

AAC is the largest group of aminoglycoside modifying enzymes, with 78 isoenzymes described and classified into three groups: AAC(2') AAC(3') and AAC(6'), which acetylate the C-2'-, C-3'- and C-6'-amino groups of 2-deoxystreptamine-containing aminoglycosides, respectively (Lovering et al., 1987). As mentioned above, *S. kanamyceticus* possesses the gene *kac* that encodes for an AAC enzyme, AAC(6')−SK (Uniprot accession: Q75PS3) that modifies kanamycin, conferring resistance to it (Murakami et al., 1983). *S. albus* encodes for ACC(6')−Isa (Uniprot accession: Q764C8) that shares 69% homology with Kac of *S. kanamyceticus*, and confers kanamycin resistance to *S. albus*, as well as to a variety of other aminoglycosides, including amikacin, dibekacin, neomycin and tobramycin (Hamano et al., 2004).

The genome of *S. leeuwenhoekii* C34 encodes for 20 putatively AAC-related proteins that harbour acetyltransferase domains usually found in AACs described in the literature (Ramirez and Tolmasky, 2010). A phylogenetic tree of the protein sequences of those 20 AAC-related proteins of *S. leeuwenhoekii* C34 and the 78 AACs described, allowed the probable classification of *S. leeuwenhoekii* C34 acetyltransferase-domain containing proteins (Figure Appendix G.4). The phylogenetic tree showed that Sle_36300 and Sle_38990 were grouped as part of the AAC−SK and AAC−Isa clade. However, they shared only 29% of identity and as *S. leeuwenhoekii* C34 was sensitive to kanamycin, it seems unlikely that those *S. leeuwenhoekii* C34 proteins might have a role in kanamycin resistance. Other groups of AACs related to kanamycin resistance are AAC(3)−IIIa, −IIIb and −IIIc (Hamano et al., 2004), but no *S. leeuwenhoekii* C34 proteins were grouped with any of those ACC representatives (Figure Appendix G.4).

In addition to these three mechanisms of resistance to kanamycin, APHs belonging to Aph(3")−I, −II, −III, −IV and −VII are a fourth kanamycin resistance determinant (Ramirez and Tolmasky, 2010; Wright and Thompson, 1999). Nevertheless, the genome of *S. leeuwenhoekii* C34 does not seem to encode for an orthologue of that group of enzymes (Figure Appendix G.3).

Apramycin is a structurally unique antibiotic that contains a bicyclic sugar moiety and

a monosubstituted deoxystreptamine that is synthesised by a specialised group of genes of *S. tenebrarius* (Walton, 1978) and *Streptoalloteichus hindustanus* (NCBI accessions: AJ629123 and AJ875019, respectively). Apramycin inhibits protein synthesis by binding to the decoding site (A-site) of the 30S subunit of the bacterial ribosome (Davies and O'Connor, 1978; Poehlsgaard and Douthwaite, 2005). Both, *S. tenebrarius* and *Streptoalloteichus hindustanus* avoid suicide by methylation of the 16S rRNA by a methyltransferase, KamB, encoded in their respective apramycin biosynthesis gene clusters (Skeggs et al., 1987). *S. leeuwenhoekii* C34 encodes for 4 putative rRNA methyltransferases (Table Appendix G.2), but none of them seemed to be homologue of any of the KamB analysed.

Another mechanism of resistance to apramycin has been associated to AACs that inactivates apramycin by acetylation. AAC(1) and AAC(3)-IV acetylate apramycin at C-1 and C-3 amino groups of apramycin, respectively (Vakulenko and Mobashery, 2003; Lovering et al., 1987). The nucleotide sequence of *aac(1)* has not been sequenced yet, so it was not possible to search for a possible homologue in *S. leeuwenhoekii* C34. AAC(3)-IV acetylates a wide range of aminoglycoside antibiotics comprising, gentamicin, tobramycin, netilmicin, sisomicin, dibekacin and apramycin (Vakulenko and Mobashery, 2003). A phylogenetic tree of AACs showed that none of the *S. leeuwenhoekii* C34 proteins was grouped together with the representative of AAC(3)-IV (Figure Appendix G.4), which suggests that *S. leeuwenhoekii* C34 does not encode for such a function in its genome, supported by observed apramycin sensitivity results.

Nalidixic acid is a synthetic quinolone antibiotic that mainly affects Gram-negative and, to a lesser extent, some Gram-positive bacteria. Its target in the cell is DNA gyrase and topoisomerase IV, both essentials for maintaining DNA topology by balancing negative superhelicity of circular DNA molecules (Huang et al., 2013). Quinolones in general, trap these enzymes as a drug-enzyme-DNA complex, blocking DNA synthesis with the subsequent release of lethal double-stranded DNA breaks (Drlica and Zhao, 1997). The preferred target of quinolones in Gram-negative bacteria is the GyrA subunit of DNA gyrase while in Gram-positive bacteria is the ParC subunit of the topoisomerase IV (Jacoby, 2005; Drlica and Zhao, 1997). Since *Streptomyces* species are naturally resistant to this antibiotic (Kieser et al., 2000), mutations might be present in these enzymes, however, bioinformatic analysis, looking for mutations in GyrA or ParC, did not lead to a conclusion.

1.5.3 Presence of divalent chloride salts ($MgCl_2$ and $CaCl_2$) improved conjugation performance between *E. coli* ET12567/pUZ8002, carrying pSET152, and *S. leeuwenhoekii* C34

Plasmid transfer by conjugation has been proven as a successful strategy to manipulate streptomycetes (Kieser et al., 2000). The success of the conjugation process between *E. coli* and *Streptomyces* is measured as conjugation frequency that is calculated as the number of *S. leeuwenhoekii* C34 exconjugants per number of donor cells, thus as higher the conjugation frequency the better the process. There are several parameters that influence the conjugation performance: i) temperature for heat-shock of spores and time

for cooling-down; ii) ratio between recipient cells and donor cells; iii) solid medium used for conjugation and salts supplemented to the medium (Du et al., 2012). The establishment of these parameters might vary from one microorganism to another, so there is no standard protocol for conjugation.

In this work, the first reported genetic modification of *S. leeuwenhoekii* C34 was achieved by conjugation with *E. coli* ET12567/pUZ8002/pSET152, using the following conjugation parameters: i) 50 °C for heat-shock of the spores; ii) the ratio between recipient cells and donor cells was $1.4 \cdot 10^7$ *S. leeuwenhoekii* C34 spores (recipient) and $1.6 \cdot 10^8$ *E. coli* (donor); iii) SFM agar supplemented with 10 mM $MgCl_2$ was used as conjugation medium. With these parameters, the conjugation frequency achieved was $1.7 \cdot 10^{-7}$, a value 170 times higher than that obtained with *S. netropsis* SD-07 ($1 \cdot 10^{-9}$) under similar conjugation conditions (Wang and Jin, 2014). This confirms that the conjugation procedures established for one actinomycete do not guarantee the same result with other actinomycetes.

Recent research has demonstrated that supplementation of divalent chloride salts to the conjugation medium has a positive effect on the conjugation performance (Yu and Tao, 2010), becoming a typical parameter that is optimised to improve conjugation frequency, although the underlying mechanism that explains this improvement is not clear (Choi et al., 2004). The concentration of $MgCl_2$ supplemented to the medium was evaluated as an initial parameter to optimise the conjugation process. In the case of conjugation between *S. licolnensis* and *E. coli* ET12567/pUZ8002/pUWL201apr, the conjugation frequency improved 2.4 times when SFM agar medium was supplemented with 20 mM $MgCl_2$ instead of 10 mM $MgCl_2$ (Du et al., 2012), while supplementation with 40 mM $MgCl_2$ instead of 10 mM $MgCl_2$ gave 7 times better conjugation frequency for *S. noursei* xinao-4 (Sun et al., 2014). In the case of *S. leeuwenhoekii* C34, the conjugation frequency was 3 times higher when SFM was supplemented with 120 mM $MgCl_2$ with respect to that obtained with 10 mM $MgCl_2$, indicating that this salt had a positive effect on the conjugation process. Conversely to our results, these authors reported that no addition of $MgCl_2$ to the medium gave no exconjugants, while we reported that no addition of $MgCl_2$ to SFM agar medium gave low conjugation frequency of $6 \cdot 10^{-8}$, implying that the salt content of soya used in this work might be contributing to the conjugation process (soya content in Appendix E.1).

Another chloride salt used in the conjugation process was $CaCl_2$. Wang et al. (2014) have shown that the presence of $CaCl_2$ had a greater effect on conjugation performance than $MgCl_2$, in fact conjugation performance improved 100 times when 60 mM $CaCl_2$ was added to the culture medium compared with $MgCl_2$ with *S. retropsis* SD-07. Also, addition of $CaCl_2$ gave higher conjugation performance with *S. coelicolor*, *S. lavendulae* and *S. venezuelae*. In the case of *S. leeuwenhoekii* C34, 60 mM $CaCl_2$ improved conjugation frequency 10 times, compared with no addition of salt, and was 5 times better than that obtained with 60 mM $MgCl_2$. Yu et al. (2010) evaluated the effect of $MgCl_2$, NaCl, $Ca(NO)_3$ and $CaCl_2$, on the conjugation frequency of *S. avermitilis* and showed that 120 mM $CaCl_2$ improved conjugation frequency over NaCl, $Ca(NO)_3$ and $MgCl_2$. Surprisingly, they showed that the combination of 60 mM $CaCl_2$ and 60 mM $MgCl_2$ improved conjuga-

tion frequency 11 fold.

Thus, the effect of the combination of $CaCl_2$ and $MgCl_2$ should be assessed for the optimisation of conjugation between *E. coli* ET12567/pUZ8002 and *S. leeuwenhoekii* C34. In addition, parameters such as the ratio between donor cells and recipient cells and the temperature of heat-shock of spores have been pointed out as other important factors that need to be evaluated (Wang and Jin, 2014). Further efforts towards optimisation of the conjugation should be aided using statistical models.

Importantly, the bactericidal effect of some antibiotics used to select genetically modified bacteria might be blocked by the presence of salts in the agar medium. Under the selected salt condition for conjugation (60 mM $CaCl_2$), only hygromycin B loses its bactericidal properties, while apramycin, kanamycin and thiostrepton were not affected (Figure Appendix G.2). Thus, the interference of the bactericidal effect of antibiotics is an indirect parameter that is necessary to consider when optimising salt concentration in the conjugation process, because the current result implicates that vector pIJ10257 (that carries a hygromycin resistance gene) might not be used for conjugation using 60 mM $CaCl_2$ in the medium.

1.5.4 Toolbox of molecular methods for *S. leeuwenhoekii* C34

S. leeuwenhoekii C34 is a novel actinomycete that produces new antibiotics, the chaxamycins and chaxalactins, whose biosynthetic routes are unknown. To study those routes in the native strain, the availability of molecular tools like cloning or integrative vectors and constitutive promoters is very important (Medema et al., 2011). In this work, the usability of several molecular tools was assessed in *S. leeuwenhoekii* C34 for further study of the biosynthesis of chaxamycins in the native strain. In addition, standard procedures for the genetic manipulation of *S. leeuwenhoekii* C34 were developed and will be discussed as follows.

One of the most commonly used strategies to genetically manipulate *Streptomyces* is by intergenic conjugation with *E. coli* (Kieser et al., 2000). In this process, a methylation-system deficient *E. coli* strain is required, since the majority of streptomycetes possess restriction systems that provide protection against foreign DNA like that coming from *E. coli* with active Dam-, Dcm- or Hsd-methylation systems. In this work an *E. coli* strain ET12567 was used since it has these three methylation systems deleted in its genome and, therefore, is suitable for replicating plasmids free of Dam-, Dcm- and Hsd-methylation.

Bioinformatic analysis of the *S. leeuwenhoekii* C34 genome revealed the presence of at least 3 putative methylation restriction-systems, homologues to those reported for *S. coelicolor* (González-Cerón et al., 2009). One of them, Sle_04470, is a conserved hypothetical protein (NCBI accession: CQR59909) that harbours an Mrr_cat domain (Pfam accession: PF04471) typically found in type II restriction enzymes containing the characteristic (D/E)-(D/E)xK active site contained in the motif of the Mrr restriction enzyme of *E. coli*

K-12, F-E-x_{24-42}-I-(**D**/E)-x_{12-15}-(**Q**/E/S/A)-x-**K**-R-x_{6-13}-(**D**/**E**) (NCBI accession: NP_418771), that recognises and cleaves DNA modified with mA or mC (Bujnicki and Rychlewski, 2001). Sco4213 of *S. coelicolor* (NCBI accession: NP_628388) also contains the Mrr_cat domain and it has been proven to restrict Dam- and Hsd-modified DNA (González-Cerón et al., 2009), both Sle_04470 and Sco4213 contain the (D/E)-(D/E)xK active site (Figure Appendix G.5A), which allows formulation of the hypothesis that *S. leeuwenhoekii* C34 encodes for a putative restriction system.

Two other putative restriction enzymes found in *S. leeuwenhoekii* C34 are Sle_34180 and Sle_45620 (NCBI accession: CQR62879 and CQR64020), homologues to Sco4631 and Sco3262 (NCBI accession: NP_628792 and NP_627474), respectively. These proteins contain the catalytic HNH_c domain (CDD accession: CD00085) of *E. coli* McrA (NCBI accession: NP_415677), an HNH nuclease known to restrict DNA modified by mC. Alignment of the five protein sequences indicates that all of them possess the characteristic histidine, asparagine, histidine/asparagine catalytic triad and four cysteines which form a putative zinc finger (Figure Appendix G.5B), likely involved in stabilising the structure (Anton and Raleigh, 2004). Sco4631 has been assigned as an endonuclease that restricts Dam-modified DNA in *S. coelicolor*, while Sco3262 restricts Dam- and Hsd-modified DNA, it seems possible that the corresponding homologues found in *S. leeuwenhoekii* C34 might have a similar function, but it remains to determine whether they are active nucleases.

On the other hand, the experimental verification of the use of the high-copy temperature-sensitive vector pGM1190 in *S. leeuwenhoekii* C34 makes possible delivery of DNA constructions for the deletion of genes of interest from the genome of *S. leeuwenhoekii* C34. In addition, the sensitivity of *S. leeuwenhoekii* C34 to thiostrepton, apramycin, hygromycinB and kanamycin, makes likely the use of vectors that possess resistance genes for these antibiotics.

The verification of the presence of canonical *attB* sites for ΦC31 and ΦBT1 in the genome of *S. leeuwenhoekii* C34 is of great interest for studying either, the expression of genes driven by strong promoters or expression driven by other kinds of promoters (own promoter of a gene or inducible promoters). Vectors like pIJ12551 and pIJ10257 harbour *attP* sites with the integrase system of bacteriophages C31 and BT1, respectively. This would allow the integration of those vectors into the genome of *S. leeuwenhoekii* C34, conferring stable transformation, which do not need the presence of the selective pressure after their integration is confirmed. Moreover, those vectors contain, just upstream to their respective multi-cloning site, the sequence of the strong promoter *ermE** of *Sacc. erythrea* that allows the over-expression of genes, useful for either *in trans* genetic complementation or for metabolic engineering purposes. pSET152, on the other hand, is a cloning vector that integrates into the *attB* site of ΦC31, it is usually used for assessment of expression of certain genes using its own promoter region.

In addition, the functionality of the *ermE** promoter was successfully tested in *S. leeuwenhoekii* C34. Transformation of *S. leeuwenhoekii* C34 with pIJ10740, a derivative of

pGUS with the *ermE** promoter driving the expression of GUS, *gusA*, allowed the hydrolysis of X-Gluc in chromogenic assays, which confirms both, *ermE** is operating in *S. leeuwenhoekii* C34 and pGUS could be used for evaluating the functionality of other promoters in *S. leeuwenhoekii* C34.

Thus, the development of standard procedures to genetically modify *S. leeuwenhoekii* C34 would allow the study of the biosynthesis of chaxamycin, which is presented in the following Chapter.

1.6 Conclusions

1. A temperature of 37 °C promotes growth and sporulation of *S. leeuwenhoekii* C34 on solid medium, whereas 30 °C yields higher chaxamycin production, in *S. leeuwenhoekii* C34.

2. *S. leeuwenhoekii* C34 is sensitive to thiostrepton, hygromycin B, kanamycin and apramycin, which allows the use of vectors carrying resistance genes for those antibiotics.

3. *S. leeuwenhoekii* C34 is capable of receiving foreign DNA (pSET152) by intergenic conjugation with *E. coli* ET12567/pUZ8002.

4. Conjugation frequency is significantly increased when the conjugation medium was supplemented with 60 mM $CaCl_2$.

5. $CaCl_2$ has a greater effect than $MgCl_2$ in promoting conjugation.

6. Vectors pGM1190 and pSET152 are suitable for use with *S. leeuwenhoekii* C34 to deliver genetic modifications.

Chapter 2

Identification and heterologous expression of the chaxamycin biosynthesis gene cluster of *Streptomyces leeuwenhoekii* C34

2.1 Abstract

De novo sequencing of the genome of *S. leeuwenhoekii* C34 was achieved by combining Illumina and Pacific Biosciences sequencing technologies, resulting in the release of a high-quality sequence of the genome and two plasmids. Genome mining allowed the identification of a stretch of genomic DNA of 80.2 kb that codes for the putative chaxamycin biosynthesis gene cluster, comprised of 27 genes: seven for biosynthesis of AHBA, five for PKS subunits, two transcriptional regulators and others for tailoring enzymes. Comparison of the putative chaxamycin biosynthesis gene cluster with others devoted to producing ansamycins allowed the identification of its putative ends. A genomic library of *S. leeuwenhoekii* C34 was screened by PCR with specific primers that amplified those putative ends. One clone, pIJ12853, contained the 80.2 kb region as part of a 145 kb insert. pIJ12853 was cloned into *S. coelicolor* M1152, a heterologous host that does not produce chaxamycins, giving rise to strain M1650. Production of all chaxamycin species was detected in the supernatant of a liquid culture of M1650, confirming that genes present in pIJ12853 encode for the chaxamycin biosynthesis route. Although it cannot be ruled out that genes of the host could be taking over biosynthesis or export tasks. Deletion of the gene that codes for the AHBA synthase (*cxmK*) in *S. leeuwenhoekii* C34 gave rise to strain M1653 unable to produce chaxamycins in liquid culture. Chaxamycin production was restored when a liquid culture of M1653 was supplemented with commercial AHBA, which confirms the identification of the chaxamycin biosynthesis gene cluster in the native strain. It is possible that supplementing analogues of AHBA might lead to the production of unnatural chaxamycins in the strain M1653.

2.2 Introduction

Nowadays, *de novo* sequencing of an entire genome is more accessible, cheaper and faster than is used to be in the past (Quail et al., 2012). Since the publication of the genome of *Streptomyces coelicolor* in 2002 (Bentley et al.), the number of sequenced genomes of streptomycetes has increased due to the rapid evolution of sequencing platforms such as, Illumina MiSeq or PacBio.

Illumina MiSeq applies the sequencing-by-synthesis technology that uses fluorescently labelled reversible terminators in each dNTP, and clusters of clonally-amplified DNA sequences (up to 700 bp) attached to a surface. This enables highly accurate sequencing results with an observed error rate of 0.80% (Illumina Inc., 2010; Quail et al., 2012). PacBio is a relatively recent sequencing platform, which applies SMRT technology. Here longer DNA sequences (up to 10,000 bp) are bonded to the DNA polymerase, which is immobilized to the bottom of 50 nm-diameter wells called ZMWs and each dNTP has a different fluorophore linked to its γ-phosphate; when a labelled dNTP is incorporated into the nascent DNA strand, the fluorescence signal is emitted in the ZMW for a short period of time and recorded in real-time. PacBio sequencing technology enables the acquisition of longer reads with an observed error rate of 12.86% (Pacific Biosciences, 2015; Quail et al., 2012).

The genome of *S. leeuwenhoekii* C34 has already been sequenced using Illumina technology and assembled in 658 contigs, giving rise to a genome size of 7.86 Mb (Busarakam et al., 2014). According to Busarakam et al. (2014), *S. leeuwenhoekii* C34 encodes for a large number of biosynthesis gene clusters that potentially produce novel compounds. However, this high number of contigs becomes into a problem when it comes to identify entire biosynthetic pathways. Some regions of the genome of *Streptomyces* contain highly repetitive nucleotide sequences, especially in those encoding for PKSs and NRPSs, resulting in erroneous assembly of the genome. Longer reads achieved with PacBio sequencing would allow better assembly of *Streptomyces* genomes, especially in those problematic regions, yielding a more reliable consensus sequence. However, the use of PacBio alone is not desirable for its relatively high error rate (Quail et al., 2012). This limitation could be bypassed by correcting the PacBio assembled genome with the more accurate Illumina reads, resulting in a very reliable sequence that would allow better genome mining analysis in *S. leeuwenhoekii* C34, towards the identification of the chaxamycin and chaxalactin biosynthesis gene clusters.

Chaxamycin A−D are naphthalenic ansamycin-type polyketides that derive from the AHBA starter unit (Figure 2.1). Chaxamycin D displays, through an unknown mechanism, highly selective antimicrobial activity against MRSA strains. Chaxamycin A, on the other hand, inhibits the intrinsic ATPase activity of the human protein Hsp90α subunit involved in cancer proliferation. In terms of structure, chaxamycins differ from other naphthalenic ansamycins, such as rifamycins, naphthomycin, geldanamycin and herbimycins in lacking a methyl group at the olefinic C-16 next to the amide bond present in these molecules (Rateb et al., 2011a).

In general, the biosynthesis of ansamycin-type polyketides starts when the type I PKS is primed with the precursor molecule AHBA. A sub-gene cluster, usually located near to the PKS genes, is largely responsible for the synthesis of AHBA (Kang et al., 2012). The enzyme AHBA synthase is encoded in the core of the AHBA gene cluster and converts aminoDHS into the final product, AHBA (Floss et al., 2011). The gene that encodes for this enzyme has been proposed as a target for PCR screening analysis, aiming at identifying new microorganisms that produce ansamycin-related metabolites (Wang et al., 2013). Indeed, *S. leeuwenhoekii* C34 encodes for an AHBA synthase gene (Rateb et al., 2011a). We predict that, in addition to AHBA, chaxamycins are synthesised from 10 additional extender units (three malonyl-CoAs and seven (2*S*)-methylmalonyl-CoAs), incorporated by a modular type I PKS.

Bioinformatic tools will enable the generation of hypotheses towards the identification of the chaxamycin biosynthesis gene cluster. For example, NCBI's CDD (Marchler-Bauer et al., 2015) and antiSMASH (Blin et al., 2013) will ease the identification of conserved domains in protein sequences and the prediction of the location of biosynthesis gene clusters in the genome of *S. leeuwenhoekii* C34, respectively. Molecular tools, such as artificial chromosomes, have been developed for cloning large genomic DNA fragments that encode for biosynthesis gene clusters (Sosio et al., 2000), allowing its expression in heterologous hosts (Gomez-Escribano and Bibb, 2011). These hosts have been developed by deleting entire biosynthesis gene cluster devoted to the biosynthesis of known compounds, whose synthesis could compete with the availability of precursor molecules when a foreign gene cluster is expressed, these hosts are called Super Hosts (Gomez-Escribano and Bibb, 2014).

In addition, the scientific evidence reported on the biosynthesis of other ansamycin-type polyketides like rifamycin of *Amycolatopsis mediterranei*, saliniketals of *Salinispora arenicola*, naphthomycin of *Streptomyces* sp. CS, among others, will be used as a reference to guide the bioinformatic identification of the chaxamycin biosynthesis gene cluster in the genome of *S. leeuwenhoekii* C34. Also, *de novo* sequencing of the genome of *S. leeuwenhoekii* C34 will be presented.

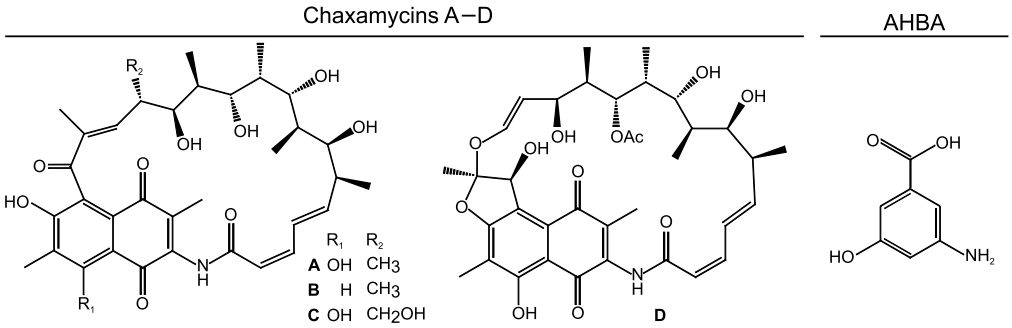


Figure 2.1: Structures of chaxamycins A−D and AHBA.

2.3 Materials and methods

2.3.1 General procedures and strains cultivation

Recipes for preparation of solid media, liquid media, solutions for microbiology and buffers can be found in Appendix E.1, Appendix E.2, Appendix E.3 and Appendix E.4, respectively.

2.3.1.1 Cultivation of *Streptomyces* and *Escherichia coli* strains

Maintenance and cultivation of *Streptomyces* and *Escherichia coli* strains was carried out according to standard protocols (Kieser et al., 2000 and Sambrook et al., 1989, respectively). Specific methods for *S. leeuwenhoekii* C34 have been largely described in Chapter One. Liquid culture of *Streptomyces* was performed in 250 ml Erlenmeyer flasks, using 50 ml of modified ISP2 medium for *S. leeuwenhoekii* C34 (chaxamycin production medium, Rateb et al., 2011a), or R3 medium for *S. coelicolor*. A seed culture of *Streptomyces* was used to inoculate fresh medium with 0.2 of initial OD measured at 600 nm (preparation in Appendix F.3.2); flasks were incubated at 30 °C and stirred at 250 rpm for 72 h. Liquid cultures of *E. coli* strains were carried out in 50 ml universal vials, using 10 ml LB supplemented with relevant antibiotics, incubated at 37 °C and stirred at 250 rpm for 20−24 h.

2.3.1.2 Analytical procedures for measurement of chaxamycins A−D

Chaxamycins A−D were detected by LC-MS/MS. A Shimadzu LC-MS system that consisted of a NexeraX2 liquid chromatograph (LC30AD) equipped with a Prominence photo diode array detector (SPD-M20A), a Kinetex XB-C18 2.6 µm 100 Å 50x2.10 mm column (part no. 00B-4496-AN, Phenomenex, USA), and a KrudKatcher Ultra HPLC in-line filter (part no. AF0-8497, Phenomenex, USA) were used for the chromatographic separation of the compounds. Then, the sample was analysed through an LCMS-IT-ToF mass spectrometer. Details about the procedure can be found in Appendix F.5. Chaxamycin A, chaxamycin B and chaxamycin C standards were kindly provided by Dr. Mostafa Rateb and Prof. Marcel Jaspars (University of Aberdeen, Scotland).

2.3.2 DNA manipulation and generation of a PAC genomic library of *S. leeuwenhoekii* C34

2.3.2.1 Sequencing of the genome of *S. leeuwenhoekii* C34

High-quality genomic DNA was isolated from *S. leeuwenhoekii* C34 using standard procedures, which are detailed in Appendix F.1.3. Sequencing of the genome of *S. leeuwenhoekii* C34 was carried out using two technologies: Illumina MiSeq (Department of Biochemistry, University of Cambridge, Cambridge, UK) and PacBio RS II SMRT (The

Genome Analysis Centre (TGAC), Norwich Research Park, Norwich, UK). In addition, the Roche 454−Junior sequencing of pIJ12853 was also incorporated into the consensus sequence. pIJ12853 contains the fully cloned chaxamycin biosynthesis gene cluster (see Section 2.3.2.2). The consensus sequence of the genome was processed with Prodigal (Hyatt et al., 2010) to identify protein-coding sequences, which then was automatically annotated with BASys (Van Domselaar et al., 2005) web server (https://www.basys.ca/). Analysis and curation of the consensus sequence was carried out with Artemis (Rutherford et al., 2000). All genome mining analyses were performed using this DNA sequence (accession number: LN831790). The full process describing the assembling and manual curation process to generate a consensus sequence has been reported elsewhere (Gomez-Escribano et al., 2015).

2.3.2.2 Genomic library of *S. leeuwenhoekii* C34 DNA and screening for the chaxamycin biosynthesis gene cluster

A PAC genomic library of *S. leeuwenhoekii* C34 DNA was constructed in pESAC13 (*E. coli-Streptomyces* artificial chromosome), which is used to create *Actinomycete* sp. libraries. *S. leeuwenhoekii* C34 genomic DNA was isolated according to the protocol described in Appendix F.1.3 and shipped to Bio S&T Inc. (Montreal, Canada) for construction and screening of the library. Briefly, high-molecular weight DNA was partly digested with *Bam*HI and ligated with pESAC13, previously digested with the same enzyme resulting in the loss of pUC19 present in the original vector (see map of pESAC13 in Appendix D.6); *E. coli* DH10B cells were used for cloning the library. The genomic library was screened for clones containing the chaxamycin biosynthesis gene cluster by PCR using two primer pairs, JFC022/JFC023 and JFC024/JFC025, designed to amplify the predicted ends of the gene cluster (Figure Appendix H.2A). One PAC clone (2-11L) was identified (Figure Appendix H.2B, C and D) and named pIJ12853. Sequencing of pIJ12853 by Roche 454−Junior sequencer (Department of Biochemistry, University of Cambridge, Cambridge, UK) confirmed the presence of the full sequence of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34.

2.3.2.3 PCR amplifications, DNA manipulation and routine sequencing

PCRs were carried out using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, USA); dNTPs stock was used at a proportion of 15:15:35:35 (A:T:G:C), since the G+C content of streptomycetes is usually 70 mol%. PCR products were cloned into pBluescript II SK(+) previously digested with *Sma*I restriction enzyme. Purification of plasmid DNA for sequencing purposes was carried out using Wizard® *plus* minipreps DNA purification system (cat. no. A7500; Promega, USA). Purification of plasmid for verification of correct cloning was carried out according to the protocol in Appendix F.1.1. Purification of DNA from the PCR mixture or agarose gel was performed with QIAquick PCR purification kit (cat. no. 28104; Qiagen, Netherlands). Eurofins Genomics service (Germany) was used for routine Sanger sequencing analysis. Primers were designed using the Primer3Plus web server (Untergasser et al., 2007). All restriction enzymes and buffers were purchased from Roche (Germany).

2.3.3 Genetic manipulation of *E. coli* and *Streptomyces* strains

List of strains, plasmids and oligonucleotides used and generated in this work are listed in Tables in Appendix A, Appendix B and Appendix C, respectively.

2.3.3.1 General procedures for genetic manipulation of *E. coli* strains

Modification of *E. coli* strains was carried out following standard procedures (Sambrook et al., 1989). Derivatives of the methylation-deficient strain *E. coli* ET12567/pUZ8002 were generated to transform *S. leeuwenhoekii* C34 or *S. coelicolor* M1152 by conjugation. The procedure for generation of electrocompetent *E. coli* ET12567/pUZ8002 can be found in Appendix F.2.2. The procedure for electrotransformation of *E. coli* is detailed in Appendix F.2.4. *E. coli* Library Efficiency® DH5α competent cells (Invitrogen, USA) were used for routine cloning and plasmid maintenance. Chemical transformation of *E. coli* DH5α was carried out following the protocol in Appendix F.2.3.

2.3.3.2 Conjugation between derivatives of *E. coli* ET12567/pUZ8002 and *Streptomyces* strains

Mobilisation of plasmids from derivatives of *E. coli* ET12567/pUZ8002 strains into *S. leeuwenhoekii* C34 was carried out following the procedure described in Appendix F.3.3. Transference of PAC DNA from *E. coli* to *S. coelicolor* was carried out according to Jones et al. (2013) and the detailed protocol is in Appendix F.3.4.

2.3.4 Construction of *S. leeuwenhoekii* strain M1653 for the deletion of AHBA synthase gene (*cxmK*)

Deletion of the AHBA synthase gene (*cxmK*) in the genome of *S. leeuwenhoekii* C34 was carried out by replacing it with the kanamycin resistance gene, *neo*, by double cross-over homologous recombination.

2.3.4.1 Construction of pIJ12851 and failure to generate *S. leeuwenhoekii* strain M1653 (Δ*cxmK::neo*)

Two DNA regions of about 1.7 kb that flank *cxmK* were cloned into the suicide vector pKC1132. These regions were amplified with primers JFC010/JFC009 (containing 5' *Hind*III and *Xba*I sites, respectively) and JFC011/JFC012 (containing 5' *Xba*I and *Bam*HI sites, respectively) and separately cloned into pKC1132, previously digested with *Hind*III plus *Xba*I and *Xba*I plus *Bam*HI to yield pIJ12857 and pIJ12858, respectively. The DNA fragment JFC011/JFC012 was excised from pIJ12858 by digestion with *Xba*I plus *Bam*HI and purified from a 1% agarose gel, then pIJ12857 was digested with the same enzymes and the fragment JFC011/JFC012 was cloned into pIJ12857, giving rise to pIJ12859. Finally, the kanamycin resistance gene, *neo*, was cloned into the *Xba*I site of pIJ12859. The

neo gene (1.3 kb) was excised from pTC192-Km by digestion with *Xba*I and purified from a 1% agarose gel and cloned into pIJ12859, previously digested with the same enzyme to yield pIJ12851 (Map of vectors and details of the procedure in Figure Appendix D.10). However, no transformants were obtained after mobilisation of pIJ12851 into *S. leeuwenhoekii* C34 by conjugation with *E. coli* ET12567/pUZ8002, therefore an alternative plasmid was used to construct *S. leeuwenhoekii* strain M1653 (Δ*cxmK::neo*) (see below).

2.3.4.2 Construction of pIJ12850 and generation of *S. leeuwenhoekii* strain M1653 (Δ*cxmK::neo*)

Two DNA regions of about 1.7 kb that flank *cxmK* were cloned into the self-replicative temperature sensitive vector, pGM1190. This vector carries the *acc(3)IV* gene that confers resistance to apramycin and which cannot replicate in *Streptomyces* at a temperature above 34°C (Muth et al., 1989) (Figure 2.2A). Construction of pIJ12850 (equivalent to pIJ12851) begun with the digestion of pIJ12859 (see previous section) with *Hind*III plus *Bam*HI to excise JFC010/JFC009-JFC011/JFC012 fragment that contains the regions that flank *cxmK*; digestion was supplemented with *Cla*I to cleave the vector backbone since this had a similar size to the excised fragment (about 3.4 kb each). The JFC010/JFC009-JFC011/JFC012 fragment was purified from a 1% agarose gel and cloned into pGM1190 previously digested with *Hind*III and *Bam*HI, giving rise to pIJ12860. Finally, the kanamycin resistance gene, *neo* (1.3 kb), was excised from pTC192-Km by digestion with *Xba*I and purified from a 1% agarose gel and cloned into pIJ12860, previously digested with the same enzyme to yield pIJ12850 (Map of vectors and details of the procedure in Figure Appendix D.11).

pIJ12850 was mobilised into *S. leeuwenhoekii* C34 by conjugation with *E. coli* ET12567/pUZ8002/pIJ12850; plates were incubated at 30 °C to ensure replication of pIJ12850 in *S. leeuwenhoekii* C34. The next day, plates were overlaid with kanamycin and nalidixic acid and incubated at 30 °C. Once exconjugants were observed, they were streaked out on SFM supplemented with kanamycin and incubated at 37 °C to promote vector loss and therefore, growing colonies under these conditions must have the *neo* gene integrated into the chromosome of *S. leeuwenhoekii* C34 by at least, one homologous recombination event. Twenty-four colonies that grew were cultivated on SFM plates for spore stock preparation and used afterwards for a screening of double cross-overs with the desired phenotype: kanamycin^R and apramycin^S (Figure 2.2A). Spores were plated on SFM agar without antibiotics and incubated in parallel at 37 °C and 30 °C. After 7 days, plates were sequentially replica-plated first on DNA agar medium supplemented with kanamycin and second on DNA agar with apramycin and incubated at 37 °C for one day. The different phenotypes were expected to be: kan^R/apra^R if pIJ12850 was still integrated in the chromosome, no second cross-over event; kan^S/apra^S if a second cross-over event excised pIJ12850 from the chromosome followed by loss of the plasmid at 37 °C or, kan^R/apra^S if a second cross-over event excised the pIJ12850 backbone (losing the apramycin resistance gene) while at the same time replacing *cxmK* with *neo* leading to the mutant Δ*cxmK::neo* (Figure 2.2B). After replica-plating, one kan^R/apra^S

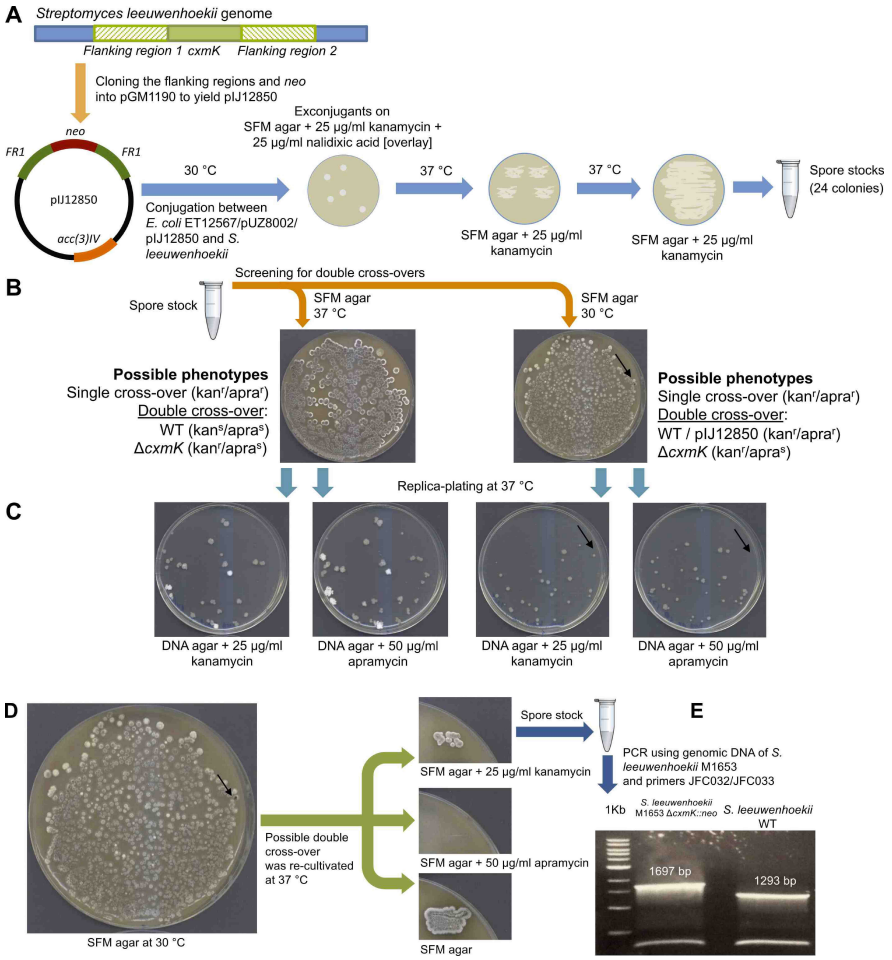


Figure 2.2: Methodology used to construct *S. leeuwenhoekii* M1653 (Δ*cxmK::neo*). Description in the main text.

colony was found (Figure 2.2C, indicated by the arrow). This candidate was streaked out sequentially on SFM agar supplemented with kanamycin, SFM agar supplemented with apramycin and SFM agar without antibiotic as control, and the $kan^R/apra^S$ phenotype expected for the Δ*cxmK::neo* mutant was confirmed (Figure 2.2D). PCR analysis with primers JFC032/JFC033 confirmed that the chosen clone had the expected amplicon size for the replacement of *cxmK* with *neo* (Figure 2.2E); the PCR product was confirmed by Sanger sequencing using the same primers; the bands obtained below 0.5 kb could be either unspecific annealing of the primers or unspecific amplification. The resulting mutant strain was called *S. leeuwenhoekii* M1653 (Δ*cxmK::neo*).

2.3.5 Complementation of *S. leeuwenhoekii* M1653 (Δ*cxmK::neo*)

2.3.5.1 Genetic complementation of strain M1653 with pIJ12852

Gene *cxmK* was PCR-amplified from the genome of *S. leeuwenhoekii* C34 with primers JFC026 and JFC034 (containing 5' *Nde*I and *Hind*III sites, respectively). The PCR product was blunt-end ligated into the pBluescript II SK(+), previously digested with *Sma*I to

yield pIJ12861. Sanger sequencing of pIJ12861 with universal primers confirmed the sequence of *cxmK*. pIJ12861 was subsequently digested with *Nde*I plus *Hind*III and the fragment containing *cxmK* was purified from a 1% agarose gel and cloned into pIJ10257, previously digested with the same restriction enzymes, giving rise to pIJ12852 with *cxmK* transcribed from the constitutive *ermE** promoter (Map in Figure Appendix D.9). pIJ12852 was mobilised into *S. leeuwenhoekii* M1653, yielding strain M1654.

2.3.5.2 Chemical complementation of strain M1653 with AHBA

AHBA was purchased from Sigma (cat. no. PH011754). Stock solution of AHBA was prepared in an Eppendorf tube by dissolving AHBA in DMSO. Liquid cultures of *S. leeuwenhoekii* C34 wild-type and M1653 were supplemented with 0.36 mM AHBA (final concentration in the flask) at 24 h after initiating the culture. After testing several concentrations and times of addition, optimal results were obtained when adding AHBA to 0.36 mM, 24 h after inoculation, presumably coinciding with the production of the chaxamycin biosynthesis enzymes and minimising AHBA degradation and/or catabolism (data not shown). Flasks were incubated at 30 °C, stirred at 250 rpm for 72 h.

2.3.6 Bioassays

Bioassays using the supernatant of liquid cultures of *Streptomyces* strains were carried out to test growth inhibition of *Bacillus subtilis* EC1524 (Institute for Food Research, Norwich, UK) and *Micrococcus luteus* ATCC 4698 as an indirect indicator of chaxamycin production.

2.3.6.1 Agar well diffusion method

General procedure for bioassays can be found in Appendix F.4. Briefly, melted LA medium was kept at 40 °C and a seed culture of either *B. subtilis* or *M. luteus* was used to inoculate the medium with 0.0125 of initial OD measured at 600 nm. The mixture was poured into a Petri dish and allowed to dry. The back of a pipette tip was used to make wells on the agar. These wells were filled with the supernatant of liquid cultures of *Streptomyces* strains; plates were refrigerated at 4 °C for 45 min and then were incubated for 1 or 2 days at 30 °C, with the surface facing upwards. A halo around a well indicates growth inhibition of the microorganism.

A variant of the described bioassay was carried out to assess the susceptibility of *S. leeuwenhoekii* C34 and *S. coelicolor* strains to chaxamycins. $1{\cdot}10^8$ spores of *S. leeuwenhoekii* C34 and *S. coelicolor* strains were spread with a cotton bud on SFM agar and LA media. The back of a pipette tip was used to make wells on the agar. One-hundred ml of supernatant of liquid culture of *S. leeuwenhoekii* C34 either wild-type (producer of chaxamycins) or M1653 (deficient in production of chaxamycins) was concentrated at reduced pressure and resuspended in 5 ml methanol to achieve a 20 times concentrated

supernatant solution. Each well was filled with 20 µl, 30 µl, 50 µl and 75 µl of this concentrated solution; 75 µl of methanol was used as control. Plates were refrigerated at 4 °C for 45 min and then were incubated for 1 or 2 days at 30 °C with the surface facing upwards. A halo around a well indicates growth inhibition of the microorganism.

2.3.6.2 Agar diffusion method with paper filter discs

$1 \cdot 10^8$ spores of *S. leeuwenhoekii* C34 and *S. coelicolor* strains were spread with a cotton bud on SFM agar and LA media. Six mm diameter paper filter discs were placed on the surface of the agar, which were hydrated with 10 µl of the supernatant of a culture of *S. leeuwenhoekii* either wild-type or M1653; modified ISP2 medium and 0.5 mg/ml apramycin were used as controls. Plates were refrigerated at 4 °C for 45 min and then were incubated for 1 or 2 days at 30 °C with the surface facing upwards. A halo around the paper disc indicates growth inhibition of the microorganism.

2.3.7 Bioinformatic analysis

Analysis of the DNA sequences obtained from sequencing results were conducted using the Staden Package (Staden et al., 1999) version 2.0.0b9 (http://staden.sourceforge.net). Searches for nucleotide or protein sequences were performed using the BLAST program (Altschul et al., 1997), at the NCBI web server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Local BLASTn or BLASTp against a database of genome/genes or proteins of *S. leeuwenhoekii* C34 was performed using NCBI package 2.2.26+ (Camacho et al., 2009) and SequenceServer (Priyam et al., in preparation) on a MacBook Pro OS X version 10.7.5. DoBISCUIT (Database of BIoSynthesis clusters CUrated and InTegrated; Ichikawa et al., 2013) was used to obtain curated and annotated versions of ansamycin-type gene clusters. Conserved protein domains were searched for in protein sequences using the NCBI's CDD Search (Marchler-Bauer et al., 2015) or the InterPro web servers (Mitchell et al., 2015). Search for conserved protein motifs was aided by ScanPROSITE web server (Sigrist et al., 2012).

2.3.8 Generation of illustrations

Illustrations of LC-MS/MS data were generated using Microsoft Office Suite. Drawings of chemical structures were done in MarvinSketch version 6.2.0. ImageJ version 1.47 (http://imagej.nih.gov/ij) and Inkscape version 0.91 (https://inkscape.org) were used to create and edit drawings and illustrations.

2.4 Results

2.4.1 The genome of *S. leeuwenhoekii* C34

De novo sequencing and assembly of the genome of *S. leeuwenhoekii* C34 gave rise to one large contig of 7,903,895 bp (7.9 Mb), which represents the chromosome (NCBI accession: LN831790.1). *S. leeuwenhoekii* C34 also carries two plasmids, the circular pSLE1 (86 kb; NCBI accession: LN831788.1) and the linear pSLE2 (132 kb; NCBI accession: LN831789.1). The sequences of the chromosome and plasmids have been deposited in the European Nucleotide Archive (PRJEB8583). Analysis of the chromosome revealed the presence of 6,712 predicted coding sequences and several potential biosynthesis gene clusters scattered along it (Razmilic et al., in preparation). In spite of the error rate of 12.86% reported for PacBio technology (Quail et al., 2012), accuracy higher than 99% was obtained for the genome sequence of *S. leeuwenhoekii* C34 when corrected the PacBio assembly with the Illumina reads. All details about the assembly and curation of the genome sequence of *S. leeuwenhoekii* C34 are beyond of the scope of this Chapter and can be found in Gomez-Escribano et al., 2015.

2.4.2 Bioinformatic identification of the chaxamycin biosynthesis gene cluster in the genome of *S. leeuwenhoekii* C34

According to Rateb et al. (2011a), chaxamycins are synthesised by a type I PKS. Our analysis indicated that the PKS that synthesises chaxamycin A consist of one Loading Module, followed by 10 extensor modules (Figure 2.3). Usually, ansamycin-type biosynthesis gene clusters span a contiguous region of the DNA that comprises genes devoted to the biosynthesis of AHBA (the precursor molecule of ansamycins), enzymes for post-synthetic modifications, regulatory proteins, antibiotic exporter and the PKS genes (Kang

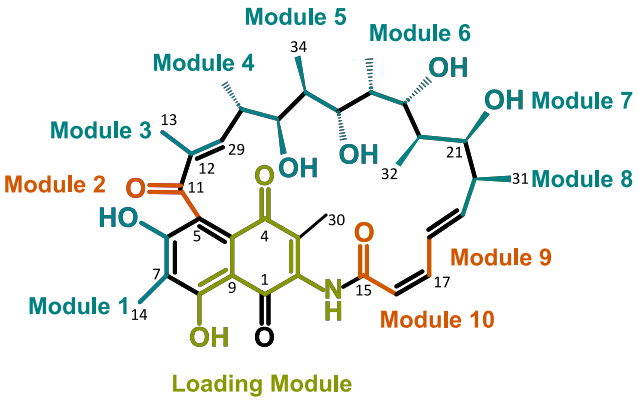


Figure 2.3: Biosynthetic origin of each extender unit present in chaxamycin A. Each extender unit was coloured and accompanied by the module that adds that extender unit: AHBA (green); malonyl-CoA (orange); (2*S*)-methylmalonyl-CoA (blue). The numbers represent the position of the carbons in the molecule (Rateb et al., 2011a). Solid wedge indicates bond or group is projecting out towards the viewer; dashed wedge indicates bond or group is projecting away from the viewer; a line indicates that the bond is in the same plane.

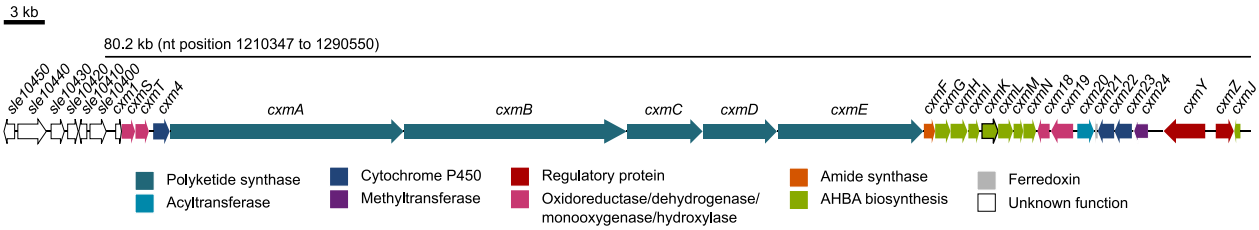


Figure 2.4: Proposed chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34.

et al., 2012). This information will be used to identify the chaxamycin biosynthesis gene cluster or *cxm* of *S. leeuwenhoekii* C34.

The enzyme AHBA synthase plays an important role in the biosynthesis of AHBA (Floss et al., 2011). The genome of *S. leeuwenhoekii* C34 contains only one gene that encodes for an AHBA synthase (*sle_10250*) whose nucleotide sequence coincides with that previously identified in this microorganism (NCBI accession FR839674.1; Rateb et al., 2011a). *sle_10250*, named *cxmK*, was found in a DNA region of the chromosome that included other seven genes devoted to the biosynthesis of AHBA (*cxmG*, *cxmH*, *cxmI*, *cxmL*, *cxmM*, *cxmN* and *cxmJ*), as well as five genes that encode for PKS subunits (*cxmA*, *cxmB*, *cxmC*, *cxmD* and *cxmE*), one amide synthase (*cxmF*), two regulatory proteins (*cxmY* and *cxmZ*), three cytochromes P450 (*cxm4*, *cxm22* and *cxm23*), one acyltransferase (*cxm20*), one methyltransferase (*cxm24*), among others (Figure 2.4). These genes are part of the proposed chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34, whose function, as well as those homologues found in the rifamycin biosynthesis gene cluster of in *A. mediterranei*, are listed in Table 2.1.

The organisation of the domains encoded in each subunit of the predicted PKS corresponds to a modular type I PKS whose number of modules is 11, consistent with the number of modules predicted for the biosynthesis of chaxamycin A (Table 2.1). The Loading Module found in *cxmA* shared high similarity with the Loading Module of other ansamycin-type PKSs that use AHBA as substrate (DoBISCUIT accession): ansamitocin (Ansam_00270; 64% identity), geldanamycin (Gelda2_00080; 68% identity), herbimycin A (Herb_00170; 69% identity), rubradirin (Rubra_00070; 70% identity) and rifamycin (Rifam_00210; 79% identity). Hence, it is likely that the Loading Module of CxmA recognises AHBA as substrate, supporting the hypothesis that this PKS is responsible for the synthesis of chaxamycins.

Table 2.1: Proposed chaxamycin biosynthesis gene cluster and protein function. Names have been assigned as far as possible according to predicted functional homology to the rifamycin gene cluster (see also Table Appendix H.1).

Sle_ number	Cxm name	Amino acids	Rifamycin homologue[1]	Proposed function in chaxamycin biosynthesis[2]
Sle_10390	Cxm1	138		Small hypothetical protein
Sle_10380	CxmS	327	RifS	NADH-dependent oxidoreductase
Sle_10370	CxmT	323	RifT	NADH-dependent dehydrogenase
Sle_10360	Cxm4	397	Orf0	Cytochrome P450

Table 2.1: *Continued. Proposed chaxamycin biosynthesis gene cluster and protein function.*

Sle_ number	Cxm name	Amino acids	Rifamycin homologue[1]	Proposed function in chaxamycin biosynthesis[2]
Sle_10350	CxmA	5616	RifA	Polyketide synthase (Module0: A_{AHBA}-kr-ACP; M1: KS-AT_{mmal}-dh-KR-ACP; M2: KS-AT_{mal}-dh-ACP; M3: KS-AT_{mmal}-kr-ACP)
Sle_10340	CxmB	5363	RifB	Polyketide synthase (M4: KS-AT_{mmal}-DH-KR-ACP; M5: KS-AT_{mmal}-dh-KR-ACP; M6: KS-AT_{mmal}-dh-KR-ACP-ACP)
Sle_10330	CxmC	1820	RifC	Polyketide synthase (M7: KS-AT_{mmal}-dh-KR-ACP)
Sle_10320	CxmD	1773	RifD	Polyketide synthase (M8: KS-AT_{mmal}-dh-KR-ACP)
Sle_10300	CxmE	3488	RifE	Polyketide synthase (M9: KS-AT_{mal}-DH-KR-ACP; M10: KS-AT_{mal}-DH-KR-ACP)
Sle_10290	CxmF[4]	269	RifF	Polyketide release and ansa-ring formation: Amide synthase
Sle_10280	CxmG[3]	368	RifG	AHBA synthesis: AminoDHQ synthase
Sle_10270	CxmH[3]	406	RifH	AHBA synthesis: AminoDAHP synthase
Sle_10260	CxmI[3]	268	RifI	AHBA synthesis: AminoQ dehydrogenase
Sle_10250	CxmK[3]	386	RifK	AHBA synthesis: Aminotransferase/AHBA synthase
Sle_10240	CxmL[3]	358	RifL	AHBA synthesis: Oxidoreductase
Sle_10230	CxmM[3]	232	RifM	AHBA synthesis: Phosphatase
Sle_10220	CxmN[3]	307	RifN	AHBA synthesis: D-kanosamine kinase
Sle_10210	Cxm18	295	Orf11	Flavin-dependent oxidoreductase
Sle_10200	Cxm19	533	Orf19	Naphthalene ring formation: FAD dependent-monooxygenase; 3-(3-hydroxylphenyl) propionate hydroxylase
Sle_10190	Cxm20	402	Orf20	Tailoring: *O*-Acyltransferase (chaxamycin D)
Sle_10180	Cxm21	63		Ferredoxin
Sle_10170	Cxm22	393	Orf4	Cytochrome P450 monooxygenase
Sle_10160	Cxm23	418	Orf5 and Orf13	Tailoring: Cytochrome P450 monooxygenase (hydroxyfuran of chaxamycin D)
Sle_10150	Cxm24	355		Tailoring: Methyltransferase, SAM-dependent (not homolog of Orf14)
Sle_10120	CxmY	433		Transcriptional regulator (C-terminus DNA-binding domain found in the NarL/FixJ response regulator family)
Sle_10110	CxmZ	246		Transcriptional regulator. Atypical response regulator of OmpR/PhoB family
Sle_10100	CxmJ[3]	168	RifJ	AHBA synthesis: AminoDHQ dehydratase

[1] Rifamycin gene cluster from *A. mediterranei* S699 (AF040570.3).
[2] Based on Pfam motif search and homology to rifamycin biosynthesis proteins. PKS domains (found with NCBI CDD Search; Marchler-Bauer et al., 2015).
[3] Genes that encode for enzymes needed for AHBA biosynthesis (Yu et al., 2001).
[4] According to Yu et al. (1999).

The delimitation of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 was defined based on synteny analysis with other related ansamycin-type biosynthesis gene clusters such as, the shared saliniketal/rifamycin biosynthesis gene cluster (*sare*) of *Sal. arenicola* CNS-205, naphthomycin biosynthesis gene cluster (*nat*) of *Streptomyces*

sp. CS and rifamycin biosynthesis gene cluster (*rif*) of *A. mediterranei* (Figure Appendix H.1; Table Appendix H.1). The right end of the chaxamycin biosynthesis gene cluster was defined as *cxmJ* that encodes for an aminoDHQ dehydratase, enzyme necessary for AHBA biosynthesis. *cxmJ* was found separated from the rest of the AHBA synthesis genes (Table 2.1, Figure 2.4), which seems to be common in other ansamycin-type biosynthesis gene clusters (Kang et al., 2012; Floss and Yu, 1999). Downstream to *cxmJ* there were no other genes of *S. leeuwenhoekii* C34 that share homology to others present in ansamycin-type biosynthesis gene clusters, nor others that could be needed to explain chaxamycin biosynthesis.

Initially, the left end of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 was defined as *cxmS* (*sle_10380*), which encodes for a putative NADH-dependent oxidoreductase, similar to those of *sare* of *Sal. arenicola* CNS-205 (*sare_1242*) and *rif* of *A. mediterranei* S699 (*rifS*) (Figure Appendix H.1). However, upstream to *cxmS* there is a region of eleven genes (*sle_10460–sle_10560*; orange genes in Figure 2.5) highly similar to those genes that participate in the biosynthesis of hopanoids, like in *S. scabies* 87.22 (*scab_12881–scab_13001*, excluding *scab_12921*; Seipke and Loria, 2009), *S. coelicolor* A3(2) (*sco6759–sco6771*, excluding *sco6761*; Poralla et al., 2000), *S. avermitilis* MA_4680 (*sav_1643–sav_1654*, excluding *sav_1647*) and in other streptomycetes.

The DNA region of 8,748 bp that is between *cxmS* and the hopanoid biosynthesis cluster encodes for seven genes, with no obvious function to include them as part of the chaxamycin biosynthesis gene cluster (white genes in Figure 2.5). Four of these seven genes encoded highly conserved proteins with homologues in *S. coelicolor*: Sle_10450, Sle_10430, Sle_10420 and Sle_10400 are homologous to Sco6772, Sco6773, Sco6774 and Sco6776, respectively. The gene *sle_10390* encodes for a hypothetical protein located immediately upstream to *cxmS*, which seems to be translationally coupled to *cxmS*; upstream to *sle_10390* there is an intergenic region of 721 bp that separates it from the next gene upstream, *sle_10400*. Therefore, *sle_10390* was selected as a probable candidate to be the left end of the chaxamycin biosynthesis gene cluster and named *cxm1*.

Thus, the proposed chaxamycin biosynthesis gene cluster spans a region of 80.2 kb, from nt position 1,210,347 to 1,290,550, and comprises 27 genes.

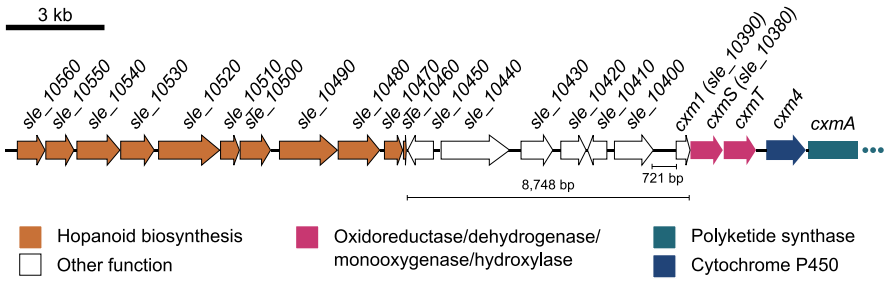


Figure 2.5: Identification of the left end of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34.

2.4.3 Analysis of reading frame-shifts in the nucleotide sequence of the chaxamycin biosynthesis gene cluster

Previously, the genome region that encodes for a PKS subunit, *cxmD*, was split into two coding sequences that showed an unusual architecture of the PKS domains. Surprisingly, there was a short DNA region of 1,162 bp in *cxmD*, whose sequence was only obtained by means of PacBio sequencing and not by Illumina nor 454 sequencing. The analysis of the GC Frame Plot (Bibb et al., 1984) of this DNA region in Artemis (Rutherford et al., 2000) suggested that there was at least one frame-shift within it. Primer pair JFC036/JFC037 were used for the amplification and Sanger sequencing of this region (covering nt position 1,243,010 to 1,244,712), whose result indicated that two guanines were not present in the PacBio sequence (positions 1,244,233 and 1,244,266). The incorporation of these two nucleotides into the consensus sequence solved the frame-shift, giving rise to one open reading frame, *cxmD* (Table 2.1). The missing guanines were located in two regions that contained four consecutive guanines and, in fact, PacBio tends to omit sequence of a C or G in homopolymeric runs of three or more Cs or Gs (Gomez-Escribano et al., 2015). Here the problem was clearly due to sequencing errors.

The nucleotide sequence of the chaxamycin biosynthesis gene cluster encodes for two fragments of putative type II thioesterase proteins that were initially named CxmR (Sle_10140) and CxmR2 (Sle_10130) due to their similarity with portions of the type II thioesterase of *A. mediterranei* S699, RifR (Figure 2.6A). RifR is a 259 amino acid type II thioesterase that is encoded in the *rif* biosynthesis gene cluster, and whose function is to preferentially hydrolyse aberrant polyketide intermediates over natural polyketide building blocks. Knock-out of the *rifR* gene resulted in a significant decrease in the rifamycin production yield, but it is not essential for rifamycin biosynthesis (Claxton et al., 2009).

Comparison of RifR with CxmR and CxmR2 indicated that the catalytic residue histidine was only present in CxmR2, whereas the catalytic residues serine and aspartate were absent in both, CxmR and CxmR2, supporting the hypothesis that they are fragments of a type II thioesterase (Figure 2.6A). *cxmR* and *cxmR2* are located between *cxm24* and *cxmY* (nt position 1,214,234 to 1,215,238), region that also encodes for two other parts of a type II thioesterase (Figure 2.6B). The amplification and Sanger sequencing of this DNA region with primers JFC053/JFC054 demonstrated that the sequence obtained matched perfectly with the published genome of *S. leeuwenhoekii* C34. This strongly suggests that possible errors in the PacBio-Illumina assembled DNA sequence can certainly be ruled out and that there exist frame-shift mutations, that resulted in a non-functional type II thioesterase. The genome of *S. leeuwenhoekii* C34 encodes for a homologue to RifR elsewhere in the genome: Sle_22920 (NCBI accession: CQR61753), a cadicidin biosynthesis thioesterase that contains the full esterase lipase domain (GrsT CDD accession: COG3208), and shares 56% identity with RifR. However, whether this or another enzyme plays an editing role in chaxamycin polyketide synthesis remains to be determined.

This part of the study gave rise to a manually edited sequence of the chaxamycin biosynthesis gene cluster that allowed all further bioinformatic analysis.

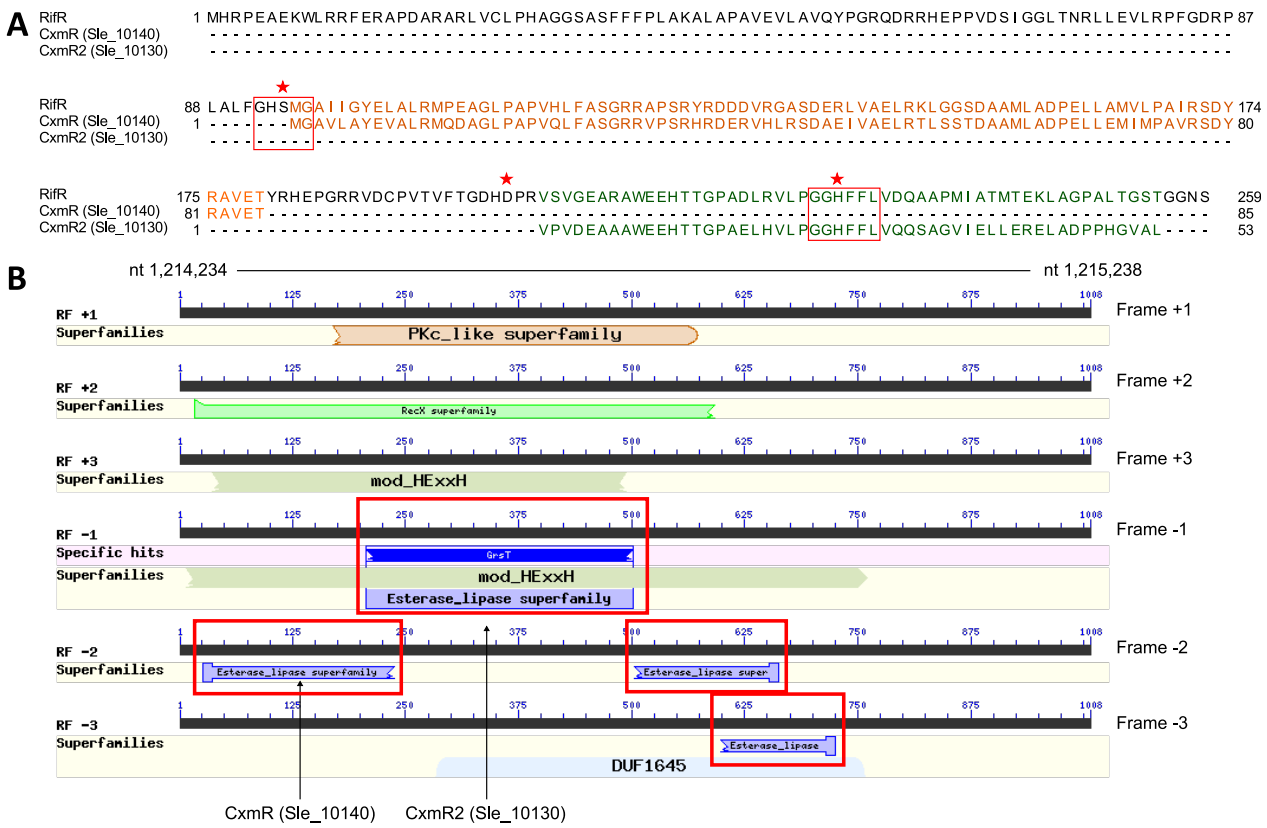


Figure 2.6: Identification of esterase lipase domains present in a region of the chaxamycin biosynthesis gene cluster that encodes for a broken type II thioesterase. (**A**) Alignment of two fragments of putative type II thioesterases found in the chaxamycin biosynthesis gene cluster: CxmR (Sle_10140, 85 amino acids; NCBI accession: CQR60477) and CxmR2 (Sle_10130, 53 amino acids; NCBI accession: CQR60476). RifR, a functional type II thioesterase of *A. mediterranei* S699 was used as control (NCBI accession: AAG52991; Claxton et al., 2009). Amino acids in orange and in green indicate the regions where CxmR and CxmR2 align within the sequence of RifR, respectively. The serine (in motif Gx**S**xG), histidine (in motif GG**H**F(F/Y)L) and aspartate residues of the catalytic triad are indicated with red stars (Claxton et al., 2009). Conserved motifs are shown contained in red boxes. (**B**) Identification of the esterase lipase domain (GrsT; CDD accession: COG3208) encoded within nt position 1,214,234 to 1,215,238. Portions of the GrsT domain, including CxmR and CxmR2 (indicated by the arrows), are in different reading frames and contained in red boxes.

2.4.4 Bioinformatic analysis of the PKS of the chaxamycin biosynthesis gene cluster

Bioinformatic analysis of the amino acid sequence of the proposed chaxamycin PKS was carried out to correlate its predicted domain architecture with the chemical structure of chaxamycin A (Figure 2.3), as a final validation. The prediction of the domains with CDD Search at NCBI (Marchler-Bauer et al., 2015) suggested that the PKS of the chaxamycin biosynthesis gene cluster comprises 11 modules, encoded in five modular type I PKS genes that synthesise the chaxamycin polyketide chain in a co-linear fashion, which means that the polyketide is elongated in series, module after module (Figure 2.7).

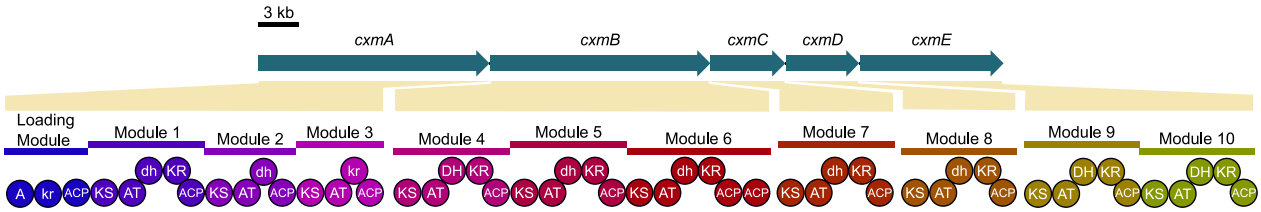


Figure 2.7: Domain architecture of the PKS found in the putative chaxamycin biosynthesis gene cluster. Conserved domains were predicted with the CDD Search at NCBI (Marchler-Bauer et al., 2015). Domain written in lower case indicates that the domain should be inactive.

2.4.4.1 Sequence-based assessment of the functionality of domains present in the chaxamycin PKS

ACP domains harbour the 4'-phosphopantetheine prosthetic group that shuttles extender units and polyketide intermediates from one module to the following one. A serine residue harbours the 4'-phosphopantetheine prosthetic group attachment site and the presence of this serine residue is indicative of normal functionality of ACP domains. Each module of the chaxamycin PKS contains an ACP domain and have the conserved residue serine, indicating that they would be active domains (Figure Appendix H.3). Interestingly, Module 6 of CxmB harbours two consecutive ACP domains in the C-terminus of that module (Figure 2.7). This is not a common feature of bacterial type I PKS, and analysis of the genome sequence indicated that the second ACP domain is not a product of a sequencing error.

AT domains select the appropriate acyl-CoA substrate and catalyse its loading onto the ACP domain. Substrate specificity of an AT domain can be predicted based on the presence of signatures in its amino acid sequence, for example AT domains that recognise malonyl-CoA as substrate contain motifs GH**S**(I/V)G and HAFH while those that recognise (2*S*)-methylmalonyl-CoA contain motifs GH**S**QG and YASH. Also, presence of the catalytic residues serine and histidine in the first and second motif, respectively, are diagnostic that that AT domain would be active (Keatinge-Clay, 2012).

The chaxamycin PKS possesses 11 modules. One of them, the Loading Module, contains an A domain that is commonly found in NRPS. This A domain would be involved in the recognition of the AHBA starting molecule, since alignment with homologues to this domain, also present in the Loading Module of other ansamycin-type biosynthesis gene clusters that recognise AHBA as substrate, revealed a high percentage of identity and the presence of motifs usually found in A domains, like the AMP-binding motif (PROSITE accession: PS00455) and the conserved ATPase motif composed by residues threonine-glycine-aspartate (Huang et al., 2001) (alignment in Figure Appendix H.7).

Modules 1 to 10 contain one AT domain each. All of them possess the catalytic residues serine and histidine in both motifs, as evidence of functionality (Table 2.2). Modules 1 and Modules 3 to 8 contain the GHSQG and YASH motifs, suggesting that they recognise (2*S*)-methylmalonyl-CoA as substrate, while Modules 2, 9 and 10 contain the GHSVG and HAFH motifs, indicating substrate preference for malonyl-CoA. The substrate specificity predicted for each AT domain was consistent with the actual extender molecule that

Table 2.2: Prediction of substrate specificity of AT domains in the chaxamycin PKS.

AT module	Signature I*	Signature II*	Predicted substrate specificity	Extender unit incorporated in chaxamycin A molecule
1	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
2	GH**S**VG	HAF**H**	Malonyl-CoA	Malonyl-CoA
3	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
4	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
5	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
6	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
7	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
8	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
9	GH**S**VG	HAF**H**	Malonyl-CoA	Malonyl-CoA
10	GH**S**VG	HAF**H**	Malonyl-CoA	Malonyl-CoA

If Signature I and II are GH**S**(I/V)G and HAF**H**, then AT recognises malonyl-CoA as substrate; if Signature I and II are GH**S**QG and YAS**H**, then AT recognise (2*S*)-methylmalonyl-CoA as substrate. Catalytic residues are in bold.
* Based on Keatinge-Clay, 2012 (Keatinge-Clay, 2012).

is incorporated by each module of the chaxamycin PKS, according to the structure of chaxamycin A (Table 2.2).

KS domains catalyse the extension of the polyketide chain through the formation of carbon-carbon bonds between an extender unit and a polyketide intermediate by means of Claisen condensation (Keatinge-Clay, 2012). The presence of the residue cysteine in the conserved motif TA**C**SSS and histidine in both EA**H**GTG and KSNIG**H**T motifs indicate that a KS domain is functional (Keatinge-Clay, 2012; Fernández-Moreno et al., 1992). Alignment of the amino acid sequences of KS domains in the chaxamycin PKS indicated that all KS domains were functional (Figure H.4).

DH domains catalyse the dehydration of the β-hydroxyl group in polyketide intermediate, leaving an unsaturation in the α,β position (Keatinge-Clay, 2012). The amino acid sequence of an active DH domain contains the conserved motifs **H**xxxGxxxxP and **D**(A/V)(V/A)(A/L)(Q/H), whose respective histidine and aspartate residue, are part of the active site (Valenzano et al., 2010). The structure of chaxamycin A contains double bonds at positions C-12−C-29, C-19−C18 and C-17−C16, indicative of dehydration and, therefore, DH of Modules 4, 9 and 10 should be active domains (Figure 2.3). Indeed, the amino acid sequence of DH of Modules 4, 9 and 10 of the chaxamycin PKS contain the conserved motif DAALH and the conserved catalytic residue histidine in the HxxxGxxxxP motif, although the DH domain of Module 9 has the proline residue substituted by alanine (Figure Appendix H.5).

The presence of both KR and DH domains in Modules 1, 5, 6, 7 and 8 might indicate that the β-ketone group of the respective polyketide intermediate would be reduced and dehydrated, with double bonds observed in the structure of chaxamycin A. However, the presence of hydroxyl groups in the structure of chaxamycin A at positions C-8, C-27, C-25, C-23 and C-21 is indicative that the DH of those Modules should be inactive domains (Figure 2.3). In fact, DH domain of Module 1 has the catalytic residue histidine mutated

to glutamine in the motif HxxxGxxxxP and lacks motif D(A/V)(V/A)(A/L)(Q/H), whereas DH domain of Module 2 lacks both conserved motifs (Figure Appendix H.5). DH domains from Modules 5 and 8, also predicted to be non-functional, are shorter than the rest of the DH domains and do not contain the motif D(A/V)(V/A)(A/L)(Q/H); thus, despite having the conserved histidine residue in the active site they lack the catalytic dyad. Surprisingly the DH domains of Modules 6 and 7 contain both active-site motifs, and we presume that there are other mutations that result in their inactivity (Figure Appendix H.5).

KR domains reduce the β-keto group in polyketide intermediates using NADPH as reducing power (Keatinge-Clay, 2012, 2007). Three diagnostic regions of the KR domains indicate whether this domain is reductase-active or reductase-inactive. Reductase-active KR domains should contain the NADPH binding site, which includes three conserved regions: a GxGxxG motif, a Swap linker and a WGxW motif (Garg et al., 2014). Alignment of the amino acid sequences of KR domains present in the chaxamycin PKS suggests that the KR of the Loading Module should be inactive, since the two second conserved glycine residue were missing in the GxGxxG motif, and it lacks the second motif WGxW (Figure Appendix H.6). In addition, KR of module 3 lacks the second conserved glycine residue in the GxGxxG motif and the glycine residue in WGxW (Figure Appendix H.6). Reductase-inactivity predicted for KR for the Loading Module and Module 3 is consistent with the structure of chaxamycin A.

Thus, sequence-based assessments of the substrate specificity and functionality of domains present in each module of the analysed PKS strongly suggest that this PKS is involved in the biosynthesis of chaxamycins in *S. leeuwenhoekii* C34.

2.4.5 Heterologous expression of the chaxamycin biosynthesis gene cluster in *S. coelicolor* strain M1152

A PAC genomic library of *S. leeuwenhoekii* C34 was built in pESAC13 and screened for the predicted ends of the proposed chaxamycin biosynthesis gene cluster. The screening of the library was carried out with primer pairs JFC022/JFC023 that amplify a fragment of 439 bp located 3.1 kb upstream to *cxm1* and JFC024/JFC025 that amplify a fragment of 340 bp located within *cxmJ*. One PAC clone was obtained after the screening, it contains the proposed chaxamycin biosynthesis gene cluster (80.2 kb) as part of a longer insert of genomic DNA of 145 kb. This clone was assigned as 2-11L and named pIJ12853 (verification of the screening process in Figure Appendix H.2).

S. coelicolor M1152 is a derivative of strain M145 that has the biosynthesis gene clusters for actinorhodin, prodiginines, type I polyketide CPK and calcium dependent antibiotic biosynthesis deleted in its genome, thus removing competition for building blocks. In addition, M1152 has a point mutation in the *rpoB*[*C1298T*] gene that confers resistance to rifampicin and that pleiotropically increases the level of secondary metabolites production (Gomez-Escribano and Bibb, 2011). *S. coelicolor* M1152 was used as a recipient of the plasmid pIJ12853 that integrates into the ΦC31 *attB* site of its genome, giving rise to

strain M1650. Liquid culture of *S. coelicolor* M1650 in R3 medium was carried out to verify whether it produces chaxamycins or not; liquid culture of M1152 in R3 medium was used as negative control while *S. leeuwenhoekii* C34 wild-type strain cultivated in modified ISP2 medium was used as positive control.

LC-MS/MS analysis of samples taken from culture of *S. coelicolor* M1650 (seventh day) demonstrates that it produced all chaxamycin species. Analysis of the EIC indicates that the retention times of chaxamycins A−D detected in the supernatant of the liquid culture of *S. coelicolor* M1650 coincide with those observed for chaxamycins A−C standards and chaxamycins A−D detected in the supernatant of liquid culture of *S. leeuwenhoekii* C34 (EIC in Figure 2.8A, C and D). As expected, no chaxamycins were detected in *S. coelicolor* M1152, the negative control (EIC in Figure 2.8B). To rule out that chaxamycin production by *S. coelicolor* M1650 was due to an unlikely cross-contamination of the culture with *S. leeuwenhoekii* C34 wild-type strain, assessment of the purity of the culture and verification of the 16S rRNA gene sequence was carried out; the results demonstrate that chaxamycins were solely produced by *S. coelicolor* M1650 (Figure Appendix H.9).

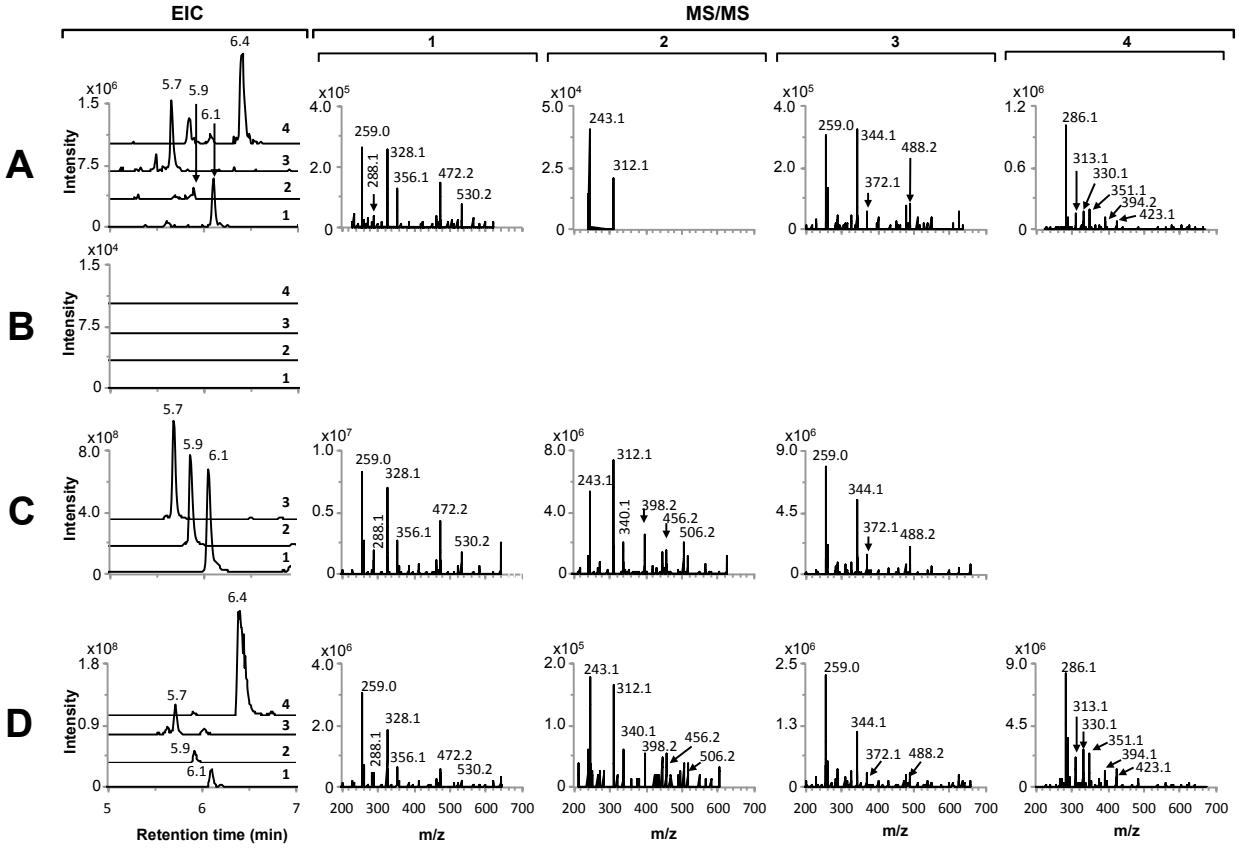


Figure 2.8: Heterologous production of chaxamycins A−D in *S. coelicolor* M1650. EIC and MS/MS fragmentation patterns are shown for each chaxamycin species detected. (**A**) *S. coelicolor* strain M1650 (M1152 containing pIJ12853); (**B**) *S. coelicolor* M1152 (negative control); (**C**) chaxamycin A, B and C standards; (**D**) *S. leeuwenhoekii* C34 wild-type strain. Chaxamycin A *m/z* 638.29 $[M - H]^-$; chaxamycin B *m/z* 622.29 $[M - H]^-$; chaxamycin C *m/z* 654.29 $[M - H]^-$; chaxamycin D *m/z* 682.29 $[M - H]^-$.

The fragmentation pattern of the chaxamycin A, C and D ions detected in liquid culture of *S. coelicolor* M1650 coincide with the fragmentation pattern of the corresponding chaxamycin A, C and D standard ions and with those found in a liquid culture of *S. leeuwenhoekii* C34 (MS/MS in Figure 2.8A, C and D). Several fragment ions were not found in the MS/MS analysis of chaxamycin B (MS/MS in Figure 2.8B), in fact only two fragment ions coincide with those found in chaxamycin B standard. Low production of chaxamycin B by *S. coelicolor* M1650 may explain this.

Chaxamycin production by *S. coelicolor* M1650 was assessed in three different media: TSB:YEME (50:50 vol%), R3 and modified ISP2 (Figure 2.9). *S. coelicolor* M1650 was able to produce chaxamycin in both R3 and TSB-YEME liquid media after 5 days of incubation at 30 °C, but no production was detected when using the *S. leeuwenhoekii* C34 chaxamycin production medium, modified ISP2. As expected, no chaxamycin production was detected in the supernatant of liquid culture of *S. coelicolor* M1152, the negative control (Figure 2.9A).

Agar well diffusion bioassays against *B. subtilis* and *M. luteus* using the supernatant of *S. coelicolor* M1650 and M1152 were carried out to correlate growth inhibition zones with the presence of chaxamycins. As expected, growth inhibition zones were readily observed with the supernatant of *S. coelicolor* M1650 cultivated in both R3 medium and TSB:YEME

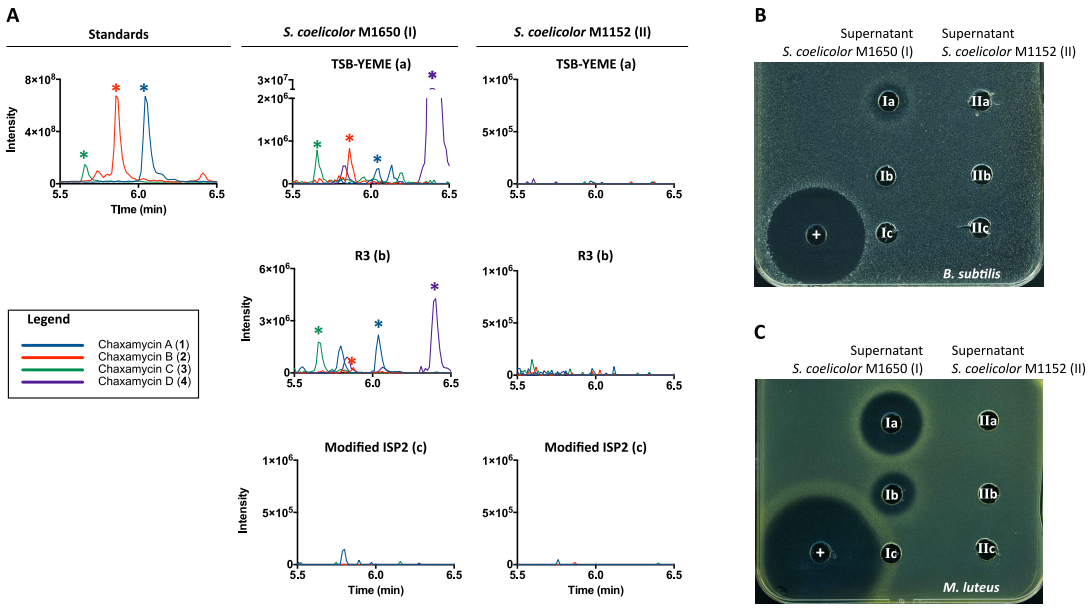


Figure 2.9: Assessment of chaxamycin production by *S. coelicolor* M1650 in different culture media and agar well diffusion bioassays against *Bacillus subtilis* and *Micrococcus luteus*. (**A**) EIC for chaxamycins A–D detected in the supernatant of liquid cultures of *S. coelicolor* M1650 (I) and *S. coelicolor* M1152 (II, negative control), cultivated in TSB-YEME (50:50 vol%) medium (a), R3 medium (b) and modified ISP2 medium (c). Only chaxamycins A–C standards were available. Supernatants were also used to assess its bioactivity against (**B**) *B. subtilis* and (**C**) *M. luteus*; carbenicillin 10 µg/ml (+) was used as positive control; growth inhibition zones of *B. subtilis* and *M. luteus* were observed only when the supernatant of *S. coelicolor* M1650 cultivated in TSB:YEME (Ia) and R3 (Ib) media was used, which is correlated with the presence of chaxamycins observed in the EICs. Chaxamycin A *m/z* 638.29 $[M - H]^-$; chaxamycin B *m/z* 622.29 $[M - H]^-$; chaxamycin C *m/z* 654.29 $[M - H]^-$; chaxamycin D *m/z* 682.29 $[M - H]^-$.

medium, in which chaxamycins were detected, while no growth inhibition zones were observed when using the supernatant of *S. coelicolor* M1152, negative control (Figure 2.9B and C). Zones of inhibition were bigger when using the supernatant of *S. coelicolor* M1650 cultivated in TSB-YEME. This observation could be explained because chaxamycin D has the strongest antibacterial activity among chaxamycins (Rateb et al., 2011a), and its production was 10 times higher when *S. coelicolor* M1650 was cultivated in TSB-YEME medium than in R3 medium.

2.4.6 Deletion of AHBA synthase gene (*cxmK*) of *S. leeuwenhoekii* C34 and chemical complementation

Biosynthesis of ansamycin polyketides begins when the AHBA starting molecule is incorporated into the Loading Module of the PKS (Floss et al., 2011). AHBA is synthesised by a specialised pathway that includes the enzyme AHBA synthase, which catalyses both the transamination of UDP-3-keto-α-D-glucose at an early stage of the pathway and the final aromatization of aminoDHS to AHBA, being essential for AHBA biosynthesis (Kang et al., 2012). The genome of *S. leeuwenhoekii* C34 harbours one copy of that gene, *cxmK*, which is located within the chaxamycin biosynthesis gene cluster. Its deletion in *S. leeuwenhoekii* C34 was performed to demonstrate that this gene is essential for chaxamycin biosynthesis.

The deletion of *cxmK* gene was carried out by replacement with the kanamycin resistance gene, *neo*, through double-crossover recombination, giving rise to *S. leeuwenhoekii* strain M1653. LC-MS/MS analysis of the samples taken from a culture of M1653 in modified ISP2 medium resulted in the complete abolition of chaxamycin production (Figure 2.10A). To conclude that the abolition of chaxamycin production was solely due to its inability to produce AHBA, because of the deletion of the *cxmK* gene, *in trans* genetic complementation of *S. leeuwenhoekii* M1653 with pIJ12852 was carried out, yielding strain M1654. pIJ12852 carries the gene *cxmK* with the constitutive *ermE** promoter driving its transcription. However, no chaxamycins were detected in the supernatant of a culture of *S. leeuwenhoekii* M1654 in modified ISP2 medium, therefore pIJ12852 failed to complement the Δ*cxmK* mutant (data not shown). Polar effects on the expression of other AHBA biosynthesis genes located downstream to *cxmK*, due to the insertion of *neo* gene, could explain this failure.

A second attempt to complement M1653 was by supplementing 0.36 mM of commercial AHBA to a liquid culture of this strain in modified ISP2 medium. This led to a restoration of the production of all chaxamycin species, as detected by LC-MS/MS (Figure 2.10B). Fragmentation pattern of chaxamycins A−D ions coincide with those found in the purified standards and *S. leeuwenhoekii* C34 wild-type (Figure 2.8C and D, respectively).

This result demonstrates that the *cxmK* gene is essential for the biosynthesis of AHBA and chaxamycins A−D, being a further evidence that the chaxamycin biosynthesis gene cluster was identified in *S. leeuwenhoekii* C34.

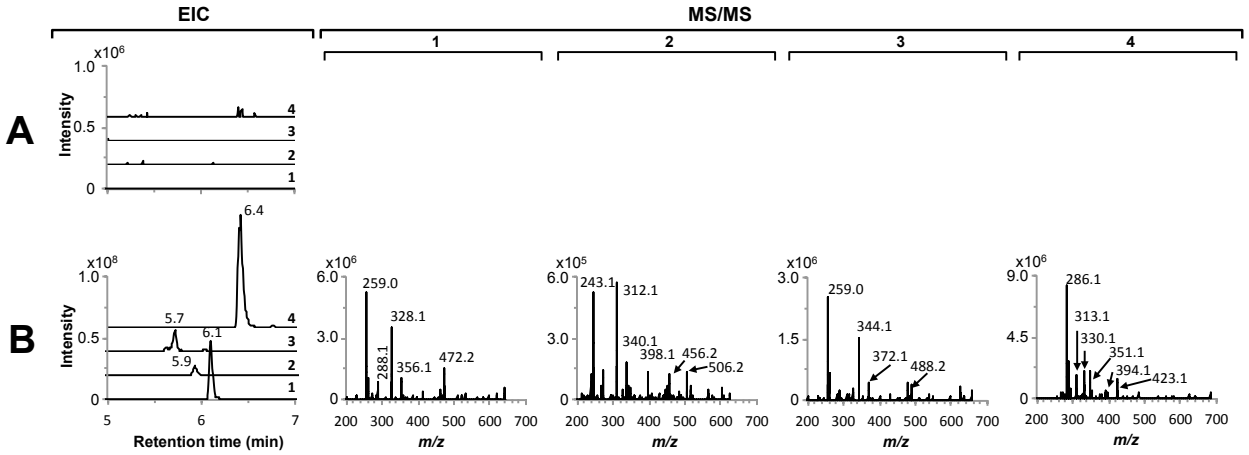


Figure 2.10: Chemical complementation of *S. leeuwenhoekii* M1653 (Δ*cxmK::neo*) with AHBA. EIC and MS/MS fragmentation patterns are shown for each chaxamycin species detected. (**A**) *S. leeuwenhoekii* M1653; (**B**) *S. leeuwenhoekii* M1653 supplemented with 0.36 mM AHBA. Chaxamycin A *m/z* 638.29 $[M - H]^-$; chaxamycin B *m/z* 622.29 $[M - H]^-$; chaxamycin C *m/z* 654.29 $[M - H]^-$; chaxamycin D *m/z* 682.29 $[M - H]^-$.

2.4.7 Assessment of sensitivity of *S. leeuwenhoekii* C34 and *S. coelicolor* strains to the presence of chaxamycins

The sensitivity of *S. leeuwenhoekii* C34 and *S. coelicolor* strains to the presence of chaxamycins was tested using both the unconcentrated and concentrated supernatants of liquid cultures of the chaxamycin producer, *S. leeuwenhoekii* C34 wild-type, and the non-producer, *S. leeuwenhoekii* M1653. Liquid cultures were used due to the low availability of chaxamycin standards.

Agar diffusion bioassays with paper filter discs were carried out using 10 µl of the unconcentrated supernatant of a liquid culture of *S. leeuwenhoekii* C34 wild-type and M1653 strains, for assessing growth inhibition of a lawn of cells of *S. leeuwenhoekii* C34 wild-type, *S. leeuwenhoekii* M1653 (Δ*cxmK*), *S. coelicolor* M1650, *S. coelicolor* M1152 and *M. luteus* (Figure 2.11). Both, *S. leeuwenhoekii* C34 wild-type and M1653 growing on LA medium presented very small growth inhibition zones when the supernatant of *S. leeuwenhoekii* C34 wild-type (that contains chaxamycins) was applied on the paper filter disc. This was not observed when the supernatant of M1653 (that does not contain chaxamycins) was used. After the third day, the diameter of those inhibition zones decreased, suggesting that the presence of chaxamycins might be exerting inhibition of the early vegetative growth that is overcame at latter growth stages. In the case of *S. coelicolor* strains, the early vegetative growth was not affected by the presence of chaxamycins. This could be explained because both *S. coelicolor* M1650 and M1152 contain a point mutation in *rpoB*[*C1298T*] gene that confers resistance to rifampicin, a compound structurally related to chaxamycins.

Growth inhibition zones of *M. luteus* were observed only when the supernatant of *S. leeuwenhoekii* C34 wild-type was applied on the paper filter disc and not with the su-

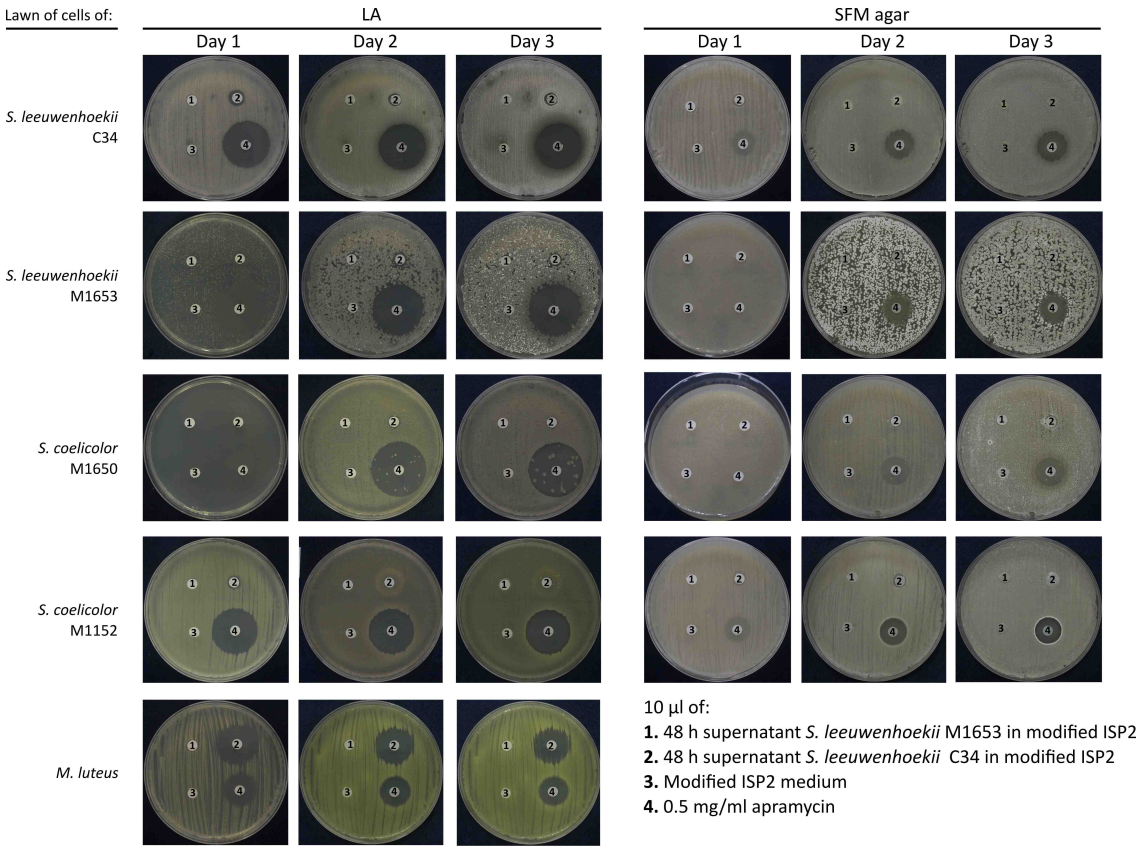


Figure 2.11: Agar diffusion bioassays with paper filter discs to test the sensitivity of *S. leeuwenhoekii* C34 and *S. coelicolor* strains to the presence of chaxamycins in supernatants of liquid culture (48 h) of *S. leeuwenhoekii* C34 wild-type and M1653. Lawns of cells were cultivated on either LA or SFM agar medium. Paper filter discs were hydrated with 10 µl of: 1 = supernatant of *S. leeuwenhoekii* M1653; 2 = supernatant of *S. leeuwenhoekii* C34 wild-type; 3 = modified ISP2 medium (negative control); 4 = 0.5 mg/ml apramycin (positive control).

pernatant of M1653, which strongly suggests that the presence of chaxamycins A–D produced those inhibition zones. However, it cannot be ruled out that the presence of other antibiotics produced by *S. leeuwenhoekii* C34, such as hygromycin A, 5"-dihydrohygromycin A, chaxalactin A–C could be contributing to the growth inhibition zones observed.

A second agar well diffusion bioassays complemented the results showed above. Different volumes of 20 times concentrated supernatant (72 h) of either *S. leeuwenhoekii* C34 wild-type or M1653 were used to assess growth inhibition of a lawn of cells of *S. leeuwenhoekii* C34, *S. leeuwenhoekii* M1653 (Δ*cxmK*), *S. coelicolor* M1650, *S. coelicolor* M1152 and *M. luteus* (Figure 2.12). In general, the size of the inhibition zones observed when using concentrated supernatant of *S. leeuwenhoekii* C34 wild-type were slightly larger compared to those when concentrated supernatant of M1653 was used (Figure 2.12A). *S. leeuwenhoekii* strains seemed to be more susceptible to inhibition by the concentrated supernatant at earlier stages (second day) compared to the fourth day of cultivation on SFM, since a decrease in the diameter of the inhibition zone was observed. *S. coelicolor* strains appeared equally sensitive to both concentrated supernatants, and therefore it is unlikely that chaxamycins are exerting inhibition on those strains, which is also supported by the fact that *S. coelicolor* M1650 is capable of producing chaxamycins. Growth

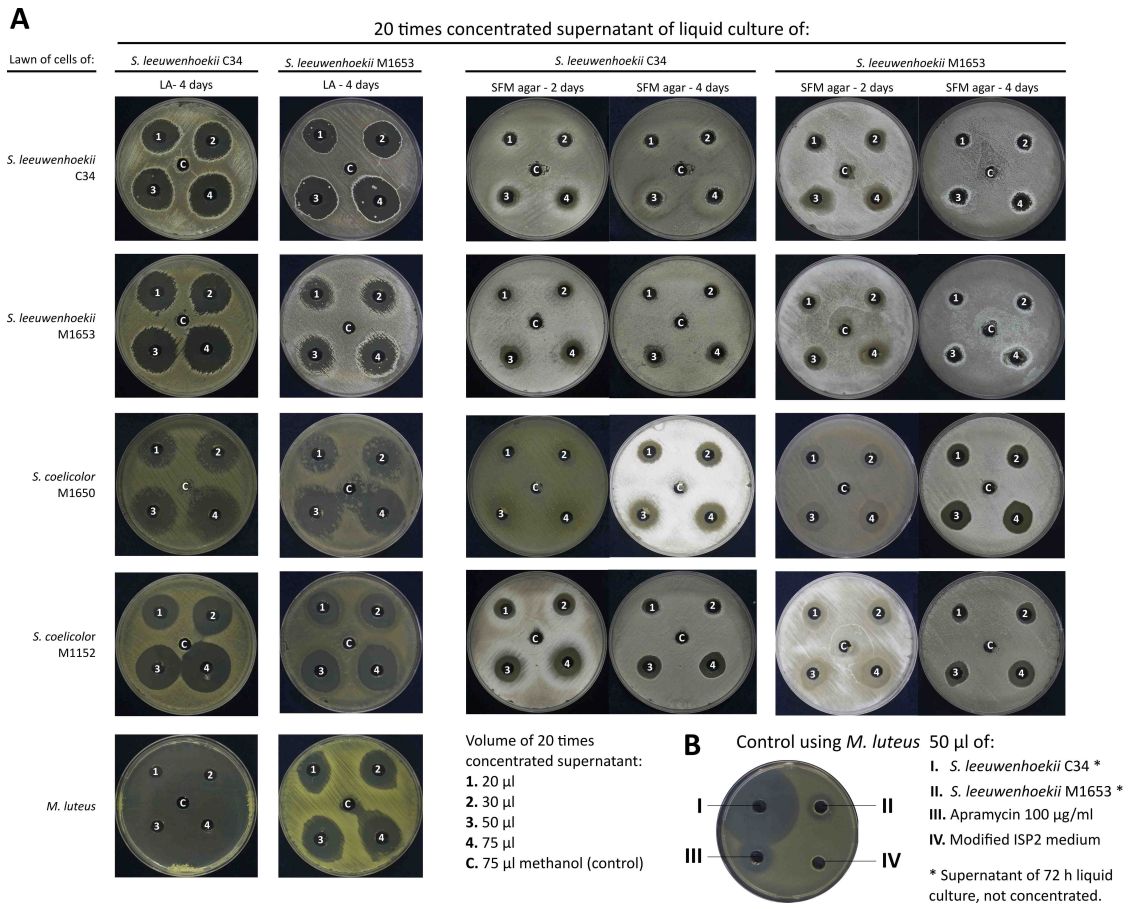


Figure 2.12: Agar well diffusion bioassay to assess of the sensitivity of *S. leeuwenhoekii* C34 and *S. coelicolor* strains to the presence of chaxamycins in concentrated supernatants of liquid culture of *S. leeuwenhoekii* C34 wild-type and M1653. Lawns of cells were cultivated on either LA or SFM agar medium. (**A**) Each well contains increasing volumes of 20 times concentrated supernatants of either *S. leeuwenhoekii* C34 wild-type or M1653 (72 h): 1 = 20 µl; 2 = 30 µl; 3 = 50 µl and 4 = 75 µl; C is a control well with 75 µl of methanol. (**B**) Bioassay with 50 µl of (unconcentrated) supernatant of liquid culture (72 h) of *S. leeuwenhoekii* C34 wild-type and M1653, not concentrated. *M. luteus* was used as indicator microorganism. 50 µl of 100 µg/ml apramycin and modified ISP2 medium were used as positive and negative controls, respectively.

inhibition zones of *M. luteus* are much larger when using concentrated supernatant of *S. leeuwenhoekii* C34 wild-type than with concentrated supernatant of M1653, which is correlated with presence of chaxamycins. Another control bioassay was performed using *M. luteus* and unconcentrated supernatants of *S. leeuwenhoekii* C34 wild-type and M1653 (the same supernatants that were subsequently concentrated). Growth of *M. luteus* was largely inhibited by the supernatant of *S. leeuwenhoekii* C34 wild-type, whereas the supernatant of M1653 inhibited growth of *M. luteus* to a lesser extent, since M1653 does not produce chaxamycin (Figure 2.12B). Other compounds that potentially could be produced by *S. leeuwenhoekii* M1653 like hygromycin A, 5"-dihydrohygromycin A, chaxalactin A–C or another unknown compound could be exerting growth inhibition of *M. luteus*.

The presence of chaxamycins does not seem to affect normal growth of *S. leeuwenhoekii* C34 and *S. coelicolor* strains in latter stages of growth. The inhibition mechanism of chaxamycins needs to be elucidated.

2.5 Discussion

The chaxamycin biosynthesis gene cluster has been bioinformatically identified, cloned and heterologously expressed in *S. coelicolor* M1152 which led to the heterologous production of all chaxamycin species. In addition, deletion of the *cxmK* gene in *S. leeuwenhoekii* C34 led to the abolition of chaxamycin production, and its restoration was achieved after feeding a liquid culture of the Δ*cxmK* mutant, *S. leeuwenhoekii* M1653, with AHBA. This information will be used to propose a putative biosynthesis pathway for chaxamycins A–D (Figure 2.13).

2.5.1 Function of the genes present in the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34

The function of the genes present in the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 have been assigned based on similarities with those homologues located in other ansamycin-type biosynthesis gene clusters such as the *rif* of *A. mediterranei*

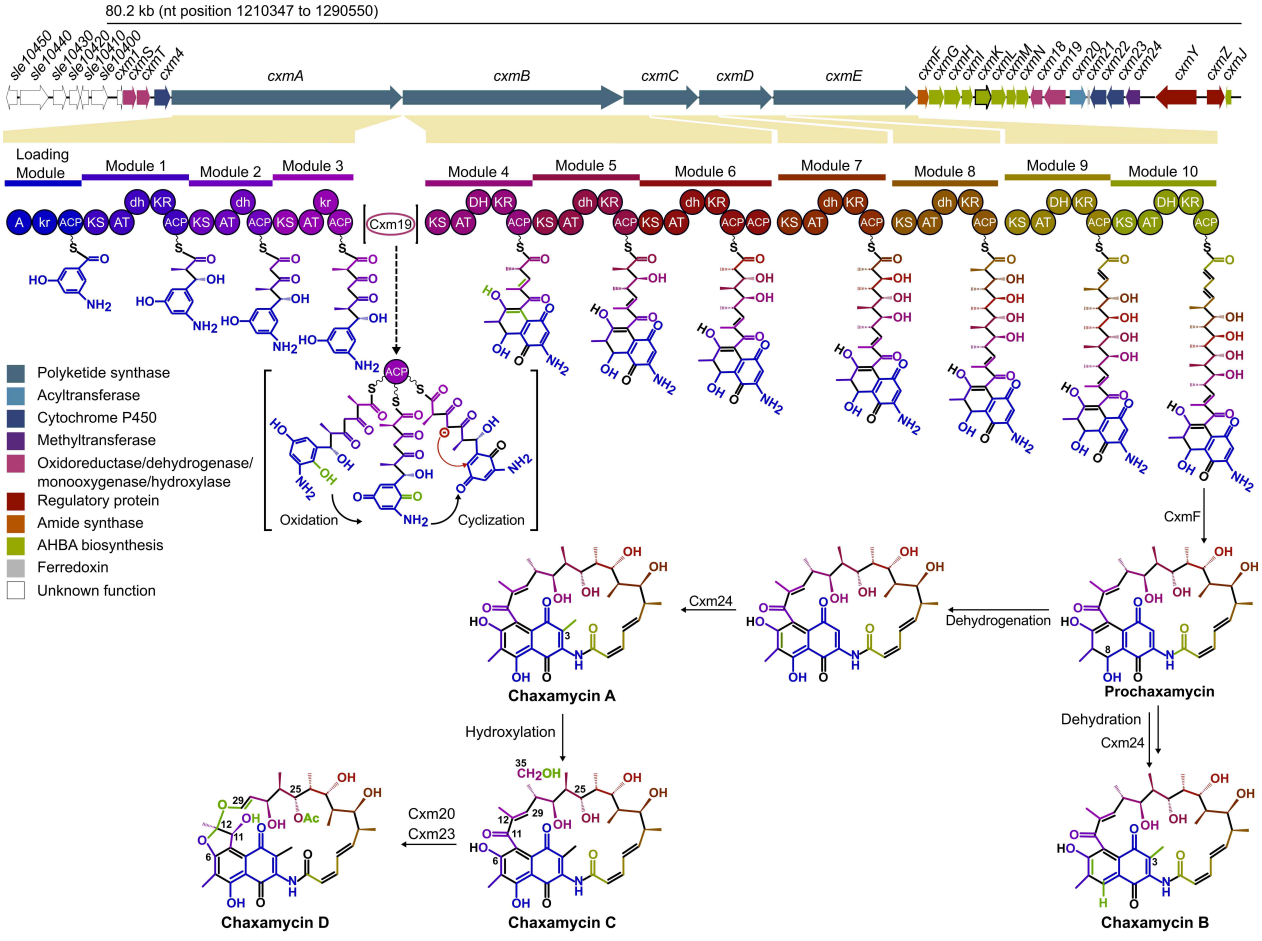


Figure 2.13: Proposed biosynthesis of chaxamycins A–D in *S. leeuwenhoekii* C34. New bonds formed are highlighted in green. Domain in lower case indicates that that domain should be inactive. Numbers in some molecules indicate position of carbon atoms.

S699, *sare* of *Sal. arenicola* CNS-205 and *nat* of *Streptomyces* sp. CS. In the following sections the biosynthesis of AHBA as well as the extension, release and cyclisation of the polyketide chain that gives rise to chaxamycins A−D will be discussed.

2.5.1.1 Biosynthesis of AHBA

AHBA is the precursor molecule of ansamycin-type polyketides (Kang et al., 2012). In *A. mediteranei*, AHBA is synthesised through the aminoshikimate pathway, which occurs parallel to the shikimate pathway (Floss et al., 2011). The chaxamycin biosynthesis gene cluster encodes for eight genes, *cxmG*, *cxmH*, *cxmI*, *cxmK*, *cxmL*, *cxmM*, *cxmN* and *cxmJ*, predicted to encode for the AHBA biosynthesis pathway. These genes share high homology and organisation with *rifG−N* and *rifJ* of *A. mediterranei* S699, which have been studied and most of them with function experimentally verified (Floss et al., 2011) (Table 2.1). This will allow the proposition of AHBA biosynthesis in *S. leeuwenhoekii* C34 (Figure 2.14).

The first step towards AHBA biosynthesis is performed by an oxidoreductase, CxmL (homologue to RifL) that catalyses the conversion of UDP-α-D-glucose into UDP-3-keto-α-D-glucose. The aminotransferase, AHBA synthase CxmK (homologue to RifK), transfers a nitrogen group from glutamine to UDP-3-keto-α-D-glucose, yielding UDP-3-amino-3-deoxy-α-D-glucose; the phosphatase, CxmM (homologue to RifM) uses UDP-3-amino-3-deoxy-α-D-glucose as substrate to give rise to D-kanosamine. D-kanosamine plays an important role in this pathway, since it carries the nitrogen atom that is incorporated in the AHBA molecule (Guo and Frost, 2002). D-kanosamine is converted into D-kanosamine-6-phosphate by the enzyme D-kanosamine kinase CxmN (homologue to RifN) (Floss et al., 2011).

Transformation of D-kanosamine-6-phosphate into 1-deoxy-1-imino-D-erythrose 4--phosphate is carried out by a transketolase, followed by an isomerase. However the chaxamycin biosynthesis gene cluster does not appear to encode for a transketolase (Table 2.1). The rifamycin biosynthesis gene cluster of *A. mediterranei* S699 possesses two genes, *rif15A* and *rif15B*, which encode for the subunits of a transketolase enzyme. A knock-out of *rif15AB* lost the ability to produce rifamycin B but accumulated rifamycin SV and S instead (Yuan et al., 2011); both rifamycin SV and S are synthesised from AHBA, which implies that other housekeeping transketolase enzymes, located elsewhere in the genome of *A. mediterranei* S699, could take over such function (Floss et al., 2011); *in trans* over-expression of *rif15AB* restored rifamycin B production in the mutant (Yuan et al., 2011). The shared rifamycin/saliniketal biosynthesis gene cluster of *Sal. arenicola* encodes for two transketolase subunits, *sare_1272* and *sare_1273*, whose deletion led to the abolition of the synthesis of AHBA (Wilson et al., 2010), suggesting that Sare_1272 and Sare_1273 are essential in the biosynthesis of that molecule.

The genome of *S. leeuwenhoekii* C34 encodes for two subunits of a transketolase, *sle_09700* and *sle_09710* that might fulfil this role. Those genes are located 47.4 kb to the right of the chaxamycin biosynthesis gene cluster, but contained within the genome

insert of 145 kb of pIJ12853 (Table Appendix H.2, genes marked in red). *sle_09700* and *sle_09710* are homologous to genes *rif15A* (74%) and *rif15B* (70%) of *A. mediterranei* S699 and to *sare_1272* (71%) and *sare_1273* (71%) of *Sal. arenicola* CNS-205, respectively. The cloning of the AHBA genes *rifG–N* and *rifJ* in other microorganisms, such as *S. coelicolor* and *E. coli* has led to the heterologous production of AHBA, which suggests that the transketolase step can be performed by an alternative enzyme encoded elsewhere in the genome since none of the cloned genes (*rifG–N* and *rifJ*) possess such transketolase activity (Yu et al., 2001; Floss et al., 2011; Yuan et al., 2011).

The compound 1-deoxy-1-imino-D-erythrose 4-phosphate is the substrate of the enzyme aminoDAHP synthase, CxmH (homologue to RifH) that produces aminoDAHP, which is then converted into aminoDHQ by CxmG (homologue to RifG). CxmJ, an aminoDHQ dehydratase (homologue to RifJ), converts aminoDHQ into aminoDHS, which is finally taken up by AHBA synthase, CxmK to yield the final product AHBA (Floss et al., 2011). AHBA is then loaded into the chaxamycin PKS to begin the extension of the polyketide chain.

CxmI is a homologue of RifI, an aminoQ dehydrogenase that catalyses the conversion of aminoDHQ to aminoQ, with no obvious role in AHBA biosynthesis. Deletion of *rifI* in *A. mediterranei* led to the an increase of AHBA production, from 350-400 mg/l to 450-500 mg/l, thus a salvage function has been proposed for the enzyme (August et al., 1998; Yu et al., 2001).

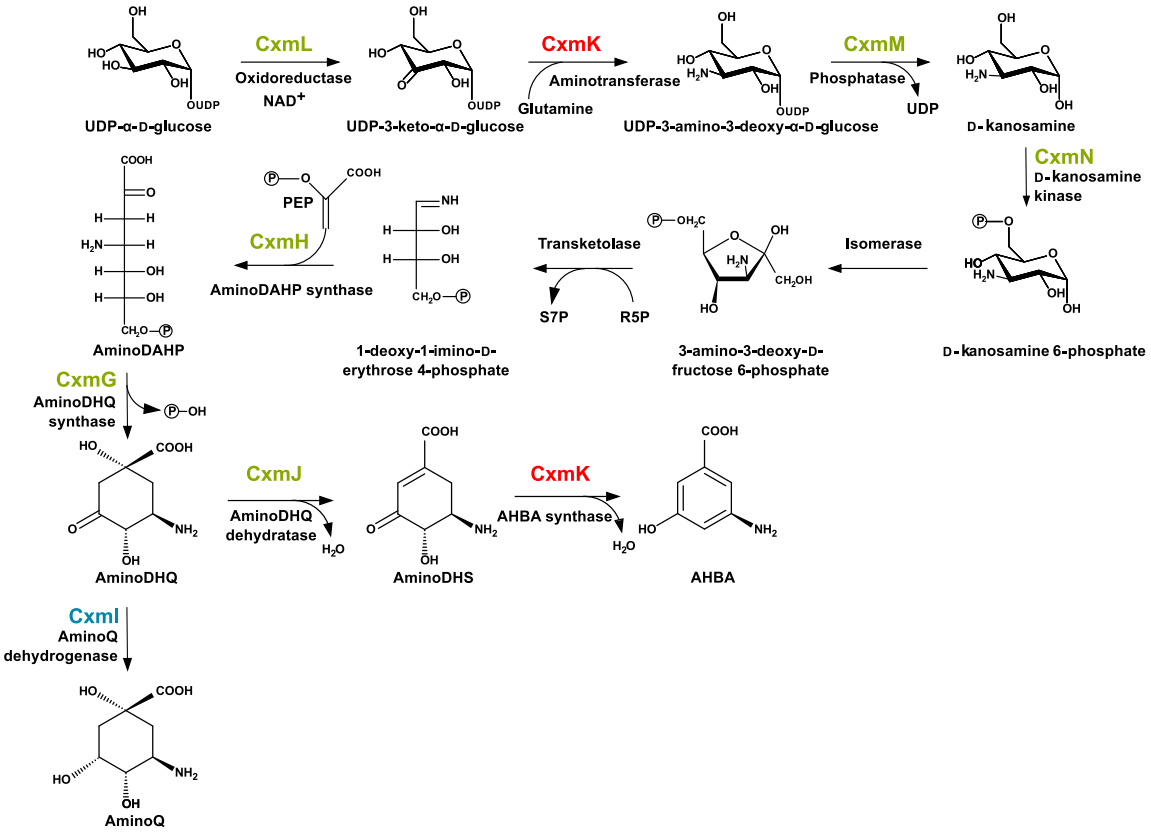


Figure 2.14: Proposed biosynthesis of AHBA in *S. leeuwenhoekii* C34. Information based on similarities with AHBA biosynthesis of *A. mediterranei*. Phosphate group is represented by Ⓟ.

2.5.1.2 Extension of the polyketide chain by the PKS

The chaxamycin PKS is composed of five subunits: CxmA, CxmB, CxmC, CxmD and CxmE (Figure 2.7). CxmA harbours the first three modules, including the Loading Module in which the extension of the chaxamycin polyketide chain starts.

The Loading Module of CxmA is a NRPS-PKS interface that contains three protein domains: an A domain, usually found in NRPS, a KR and an ACP. Adenylation domains are commonly found in the loading module of the rapamycin PKS (Park et al., 2010) and in ansamycin-type PKS, as in RifA of *rif*, NatA of *nat*, and others that also recognise AHBA as substrate (Admiraal et al., 2001; Wu et al., 2011). Alignment of the amino acid sequence of NRPS-like A domains indicated that the AMP-binding pocket and the ATPase motifs, both essential for its functionality, are highly conserved in the A domain of CxmA (Figure Appendix H.7), suggesting that it might work similarly to its homologue in RifA (Figure 2.15).

The mechanism on how the A domain of RifA works has been described by Admiraal et al. (2001). In the rifamycin biosynthesis, the loading module of the PKS activates the unprotonated form of AHBA starting molecule, AHB, with ATP, yielding an AHB-AMP species

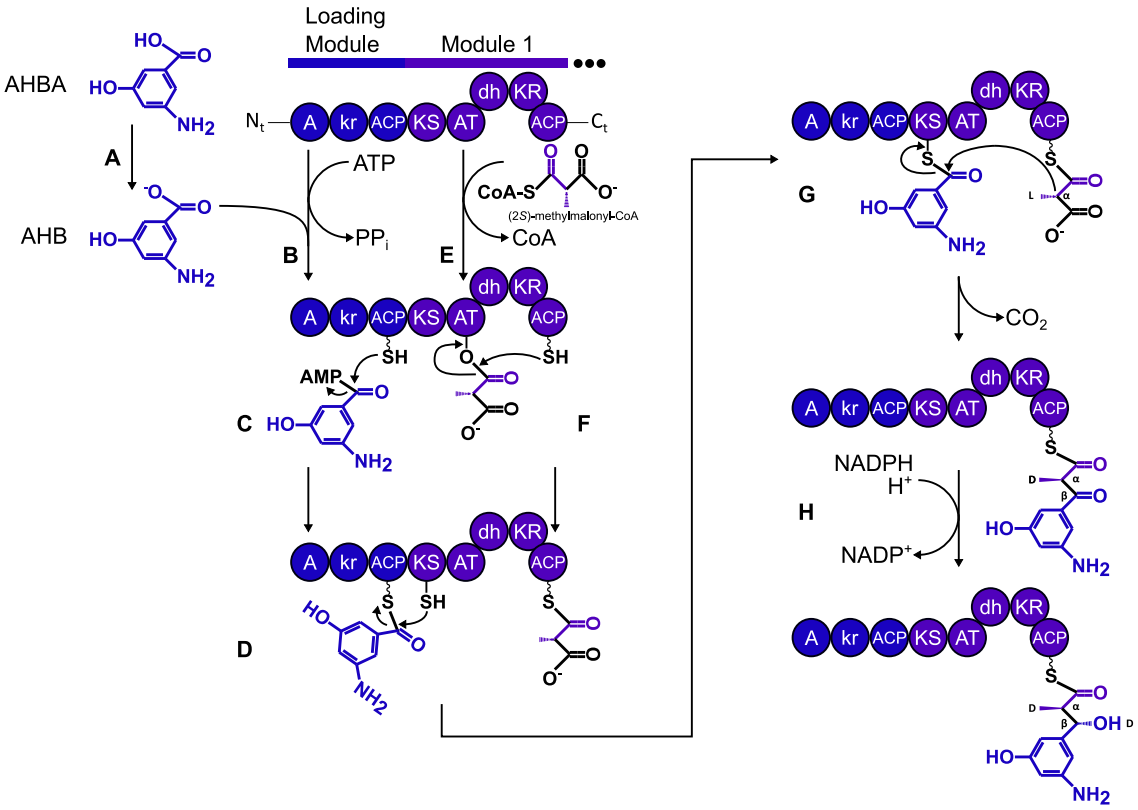


Figure 2.15: Proposed mechanism for AHBA loading into the Loading Module of the chaxamycin PKS. This Figure was adapted from the proposed mechanism for substrate loading in the Loading Module of the rifamycin PKS (Admiraal et al., 2001, 2002). For simplicity, only the Loading Module and Module 1 of the chaxamycin PKS are shown. Lower case indicates that the domain should be inactive. Details in the main text.

that is then linked to an ACP domain, located next to the A domain in the Loading Module; this mechanism is analogous to the activation and covalent attachment of amino acids in canonical NRPS modules, which is independent of coenzyme A ligase (Admiraal et al., 2001). The A domain of CxmA might carry out the activation of AHBA, similarly as it occurs in RifA; in this case, the unprotonated form of AHBA would be activated by the A domain using ATP and then the AHB-AMP species generated would be linked to the ACP domain of the Loading Module (Figure 2.15A, B and C). The KR domain of the Loading Module, also located the Loading Module of CxmA, is predicted as reductase-incompetent because the GxGxxG motif is incomplete and the WGxW motif is absent; both are essential for binding NADPH, which confer reductase activity (Figure Appendix H.6).

The extension of the polyketide chain carries on in Module 1, which is located right after the Loading Module and harbours protein domains KS, AT, DH, KR and ACP (Figure 2.13). The AT domain of Module 1 is predicted to recognise (2*S*)-methylmalonyl-CoA, which is covalently bonded to a serine located in its active site (Figure 2.15E). Then the thiol group of the 4' phosphopantetheine prosthetic group present in ACP domain of Module 1, attacks the methylmalonyl moiety, previously bonded to the AT domain, resulting in its transfer to ACP domain (Figure 2.15F). In the next step, the AHB moiety bonded to the ACP domain of the Loading Module is transferred to the KS domain of Module 1, reaction catalysed by this KS domain (Figure 2.15D→G). Claisen condensation between the methylmalonyl moiety bonded to the ACP domain of Module 1, and the AHB intermediate linked to the KS of Module 1 results in the formation of a diketide polyketide intermediate, reaction that releases CO_2 from the methylmalonyl moiety, which also inverts the orientation of the α-substituent from L to D (Keatinge-Clay, 2012) (Figure 2.15G→H). Finally, the KR domain of Module 1 catalyses the reduction of the β-ketone group, present in the AHB moiety of the diketide, resulting in a β-hydroxyl group (Figure 2.15H). Analysis of the amino acid sequence of the KR domain of Module 1 indicates that it belongs to B1−type KR (Keatinge-Clay, 2007), which implies that the final orientation of the α-methyl and the β-hydroxyl groups are D and D, respectively. However, there is no structural evidence of this configuration in the chemical structure of chaxamycin A to confirm whether this prediction is correct or not, since the stereochemistry of the α- and β- chiral centres are probably lost during the formation of the naphthalenic ring by Cxm19 (see below).

Incorporation of malonyl-CoA by Module 2 and (2*S*)-methylmalonyl-CoA by Module 3, results in the formation of a triketide and tetraketide, respectively (Figure 2.13). In the case of the triketide, the β-ketone remains intact, since Module 2 lacks a KR that could perform this task; Module 3 harbours a KR domain that was predicted as inactive, because of the absence of the tyrosine residue in the catalytic triad (Figure Appendix H.6). The formation of the tetraketide might involve a change in the orientation of the α-methyl from L to D (Keatinge-Clay, 2012); however, there is no evidence of this modification in the structure of chaxamycin A that supports this prediction, probably because of the subsequent formation of the naphthalenic ring by Cxm19 (see below).

In rifamycin biosynthesis, before the rifamycin-tetraketide enters Module 4 to yield a pentaketide, it undergoes an enzymatic reaction catalysed by Rif-Orf19, a 3-(3-hydroxyphenyl)-propionate hydroxylase, that is responsible for the formation of the

naphthoquinone ring (Xu et al., 2005). The biosynthesis of another ansamycin-type polyketide, naphthomycin, also involves the formation of a naphthoquinone ring that has been demonstrated to be carried out by Nat2, encoded in *nat* of *Streptomyces* sp. CS, which is a homologue of Rif-Orf19 (Wu et al., 2011). This process occurs while the polyketide chain is being synthesised, in between modules 3 and 4 of the respective PKS (Floss and Yu, 1999; Yu et al., 1999). A mechanism for the formation of the naphthoquinone ring in rifamycin biosynthesis has been proposed by Floss et al. (1999). In the first step, a *p*-hydroxylation of the benzene ring in the AHB moiety of the tetraketide leads to oxidation to yield a quinone, then a cyclization step occurs by Michael addition to yield the naphthoquinone ring. The chaxamycin biosynthesis gene cluster also encodes for Cxm19, a homologue of Rif-Orf19 and Nat2, that might carry out a similar process in the chaxamycin tetraketide, giving rise to a tetraketide that has the structure of a naphthoquinone ring (Figure 2.13, see step carried out by Cxm19).

In the chaxamycin biosynthesis, the KS domain of Module 4 catalyses Claisen condensation between the tetraketide modified by Cxm19 and the methylmalonyl moiety, previously loaded into the ACP of Module 4 to form a pentaketide. The KR domain of Module 4 reduces the β-ketone to hydroxyl, which is then dehydrated by DH of the same Module, leaving a double bond in the β,γ-position (Figure 2.13). The position of this unsaturation in the chaxamycin pentaketide is atypical; commonly, after the dehydration process carried out by a DH domain of a PKS, the double bond lies in the typical position between the α,β-carbons of the polyketide intermediate (Keatinge-Clay, 2012). This atypical position has also been found in other ansamycin-type polyketides such as rifamycins (Xu et al., 2005), naphthomycins (Wu et al., 2011) and ansamitocins (Taft et al., 2009). Only in the case of ansamitocins, the shift of two double-bonds from the typical α,β- to the atypical β,γ-position has been experimentally demonstrated that was carried out by the DH of module 3 of the ansamitocin PKS (Taft et al., 2009). Another example comes from the rhizoxin PKS from the endosymbiont bacteria *Burkholderia rhizoxinica* of the rice-seedling-blight fungus *Rhizopus microsporus* (Kusebauch et al., 2010). Here, the DH of module 7 of the *trans*-AT PKS introduces the unsaturation with concomitant shift of the double-bond from the α,β-position to the β,γ-position (β,γ-dehydration), which was observed in the final structure of rhizoxin. Thus, it seems that the atypical position of the double bond found in the chaxamycins could be formed by the action of the DH domain of Module 4 on the pentaketide, but it remains to be experimentally demonstrated.

The following steps towards the biosynthesis of the chaxamycin polyketide chain are sequentially carried out by Modules 5 to 8 that incorporate (2*S*)-methylmalonyl-CoA units, giving rise to intermediates from a hexaketide to a nonaketide, respectively. The KR domain of each of these Modules performs the reduction of the β-ketone group to β-hydroxyl. No dehydration reaction would be performed in those Modules, since their DH domains are predicted as inactive. The final step of the extension of the chaxamycin polyketide chain is performed in CxmE that contains Modules 9 and 10. Those Modules add malonyl-CoA units to form the decaketide and undecaketide, respectively, that contain α,β-unsaturations incorporated by KR and DH domains present in their respective Modules. Afterwards, the undecaketide is released and cyclised (Figure 2.13).

2.5.1.3 Release and cyclisation of the chaxamycin polyketide chain

In rifamycin biosynthesis it has been demonstrated that the amide synthase RifF, encoded in the gene *rifF*, is a chain-termination enzyme that performs the release of the polyketide chain with the concomitant formation of an intramolecular amide bond between the amino-group of the AHBA moiety and the carbon atom of the last malonyl moiety incorporated by Module 10, to yield a cyclic compound called proansamycin X. Proansamycin X is the base molecule for the biosynthesis of all rifamycins (Yu et al., 1999). Knock-out of *rifF* gene in *A. mediterranei* S699 results in a mutant strain that no longer synthesizes rifamycin B, but analysis of the organic extract revealed the presence of a series of linear rifamycin polyketide chains, from the pentaketide to the decaketide intermediates (Yu et al., 1999).

The chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 encodes for an amide synthase, CxmF, that shares 76% identity with RifF (NCBI accession: AAC01715.1) and with other amide synthases such as, Sare_1251 of *sare* gene cluster (72%; NCBI accession: ABV97156.1) and NatF of *nat* gene cluster (55%; NCBI accession: ADM46361). CxmF possesses a catalytic triad composed of cysteine, histidine and aspartate that is very conserved across amide synthases of different microorganisms (Figure Appendix H.8). Thus, it is proposed that CxmF catalyses the release and cyclisation of the undecaketide to form a new molecule called prochaxamycin that is thought to be the base molecule of all chaxamycin species (Figure 2.13). Although the structure of prochaxamycin has not been clearly identified through HPLC-MS of liquid cultures of *S. leeuwenhoekii* C34, probably because it would be rapidly converted into other chaxamycin species, it has been carefully deduced from the sequential biosynthesis of the chaxamycin polyketide chain. Prochaxamycin would be equivalent to pronaphthomycin, the base molecule of the naphthomycins (Wu et al., 2011) and to proansamycin X, the base molecule of the rifamycins (Yu et al., 1999). In addition, prochaxamycin enables postulation of how chaxamycins A−D are synthesised.

2.5.1.4 Biosynthesis of chaxamycin A, C and D

The biosynthesis of chaxamyicn A starts with the dehydrogenation of the prochaxamycin molecule at the C-8 position that generates a new bond in the naphthoquinone ring (Figure 2.13, see dehydrogenation reaction), which is in agreement with the structures of chaxamycin A and C. A similar dehydrogenation step has been proposed for the synthesis of rifamycin W from proansamycin X, which has been experimentally demonstrated (Yu et al., 1999).

Cxm24, a SAM-dependent methyltransferase (not homologue to Rif-Orf14), that is likely to perform the methylation of prochaxamycin at the C-3 position, which yields the first known product of this biosynthetic route, chaxamycin A (Figure 2.13, C-3-methylation in chaxamycin A is indicated). The C-methylation of the naphthoquinone chromophore is apparently unique to ansamycin biosynthesis (Kang et al., 2012) and it has only been detected in streptovaricin C, an ansamycin polyketide derived from acetate, propionate, methionine, and AHBA precursors (Staley and Rinehart, 1991). The biosynthetic origin of

this methylation has been assumed to come from methionine, which was confirmed by feeding with [methyl-^{13}C]methionine precursor. Unfortunately, the biosynthesis gene cluster of streptovaricins has not been identified and therefore comparative analysis of the putative chaxamycin methyltransferase, Cxm24, with the methyltransferase, possibly located within the streptovaricin gene cluster, is not possible. Furthermore, the *nat1* gene of the *nat* gene cluster encodes for a chlorinating enzyme that chlorinates naphthomycin E at an equivalent position with respect to the C-3-methylation observed in chaxamycins A–D, this suggests that the C-3 position of prochaxamycin (equivalent to position C-4 in AHBA molecule) could be a target for chlorination to produce chlorinated unnatural chaxamycin derivatives (Wu et al., 2011).

The biosynthesis of chaxamycin C is believed to occur by a hydroxylation in position C-35 of chaxamycin A (Figure 2.13, C-35-hydroxylation in chaxamycin C is indicated). Although the chaxamycin biosynthesis gene cluster has been identified, there is not a clear candidate that could perform this reaction. In the case of rifamycin biosynthesis, rifamycin W is formed after a series of reactions on proansamycin X, one of those involves an hydroxylation at C-34a position, which is located at an equivalent position to the C-35 hydroxylation of chaxamycin A. Xu et al. (2005) postulate that the C-34a alcohol of rifamycin W may be incorporated by a P450-dependent monooxygenases like *rif0*, however, mutational analysis of this gene did not give a conclusive result. Alternatively, they propose that this reaction could be taken over by a P450 enzyme with relaxed substrate specificity. The chaxamycin biosynthesis gene cluster encodes for several cytochromes P450 that could perform this reaction.

Chaxamycin C is proposed as the substrate for the biosynthesis of chaxamycin D. *cxm23* of the chaxamycin biosynthesis gene cluster encodes for a cytochrome P450 monooxygenase that is a homologue to *rif-orf5* of the *rif* gene cluster. Rif-Orf5 has been experimentally verified as responsible for the polyketide backbone rearrangement of rifamycin W that gives rise to DMDARS (Xu et al., 2005). DMDARS contains a hydroxyfuran ring, the same type of structure that is observed in chaxamycin D. Thus, Cxm23 was proposed as a candidate that, as its homologue Rif-Orf5, oxidatively cleavages C-12–C-29 double of chaxamycin C that leads to a rearrangement of the polyketide that forms the observed ketal structure in chaxamycin D (Figure 2.13, see rearrangement of C-6-oxygen, C-11 and C-11-ketone, C-12 and C-29 of chaxamycin C and D).

The biosynthesis of chaxamycin D finishes with an *O*-acetylation at C-25. In the biosynthesis of rifamycin B, the conversion of DMDARSV to DMRSV involves the *O*-acetylation of C-25, which is catalysed by an acetyltransferase encoded in *rif-orf20* and may use acyl-CoA substrates such as, propionyl-CoA or acetyl-CoA (Xiong et al., 2005). Cxm20, an *O*-acyltransferase, shares high homology with Rif-Orf20 and it was hypothesised to catalyse the C-25-*O*-acetylation to yield chaxamycin D (Figure 2.13, C-25-oxygen in chaxamycin D).

2.5.1.5 Biosynthesis of chaxamycin B

The biosynthesis of chaxamycin B is proposed as an alternate pathway to the biosynthesis of the rest of the chaxamycins. This pathway involves a dehydration of the C-8-hydroxyl group of prochaxamycin, followed by a C-3-methylation of the intermediate by Cxm24 to finally yield chaxamycin B. This is an analogue processes to that proposed for the conversion of pronaphthomycin into naphthomycin E in *Streptomyces* sp. CS. Pronaphthomycin, which is equivalent to proansamycin X and prochaxamycin, is thought to undergo a dehydration step to give rise to naphthomycin E, a compound that is similar to chaxamycin B. Even though no obvious gene that would encode for a dehydratase has been identified either in the *nat* gene cluster or in the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34, a dehydration step would be required for the aromatization of the naphthalene ring to generate naphthomycin E (Wu et al., 2011); a similar process is thought to occur in the biosynthesis of chaxamycin B.

2.5.2 Prediction of the stereocontrol applied by KR domains over the chaxamycin polyketide intermediates

When a KR domain reduces the β-ketone group of a polyketide intermediate, it exerts stereocontrol on both the α-substituent (if present) and β-hydroxyl groups of that intermediate and, if there is no other modification (dehydration, for example), the stereocontrol applied by the KR is likely to remain in the final structure of the molecule synthesised. Bioinformatic studies on the amino acid sequences of KR domains of known stereocontrol have shown that the presence or absence of certain fingerprints allow the prediction of the likely stereocontrol exerted by KR domains over the α-substituent and β-hydroxyl groups of a polyketide intermediate (Keatinge-Clay, 2007). In the case of chaxamycin A, stereocontrol exerted by KR domains located in Modules 4 to 8 can be determined analysing substituent groups bonded to C-20 to C-29 (Figure 2.3).

According to Keating-Clay (2012), KR domains can be classified as A−, B− or C−type. A−type KR domains generate a β-hydroxyl group with L-orientation (*S* configuration for the chiral β-carbon); those KR domains have a conserved a tryptophan residue to exert this type of stereocontrol. B−type KR domains generate a β-hydroxyl group with D-orientation (*R* configuration for the chiral β-carbon); they have a conserved leucine-aspartate-aspartate (LDD) motif, the leucine is occasionally replaced by valine or isoleucine, the first aspartate can be replaced by several residues but, the second aspartate is strictly conserved. C−type KR domains are reductase-incompetent since they lack the active tyrosine residue. If a KR domain operates on an α-substituted polyketide, it can be further classified based on the orientation of the α-substituent of the product of the reduction: α-substituent in D-orientation, then *1* is appended; α-substituent in L-orientation, then *2* is appended, which yields A1−, A2−, B1− or B2−type KR domains that operate on α-substituted intermediates (Figure 2.16).

Importantly, if the intermediate is immediately processed by an active DH present in the

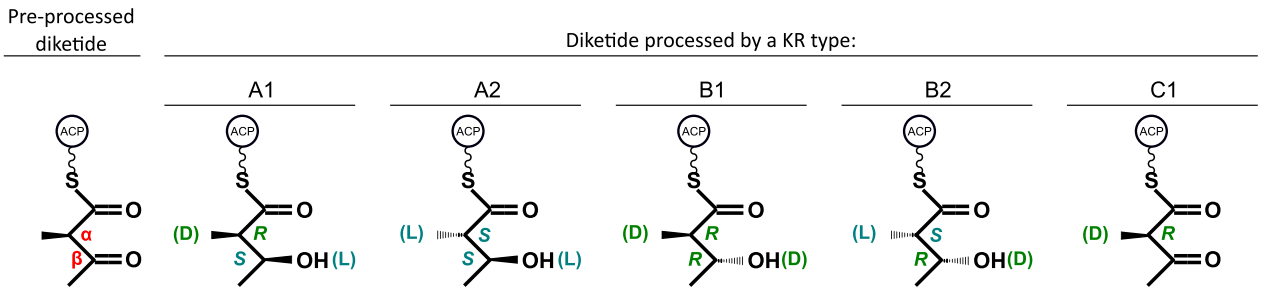


Figure 2.16: Classification of KR domains of PKS according to the stereocontrol exerted on the β-ketone group and α-substituent of a polyketide intermediate. Figure adapted from (Keatinge-Clay, 2007, 2012).

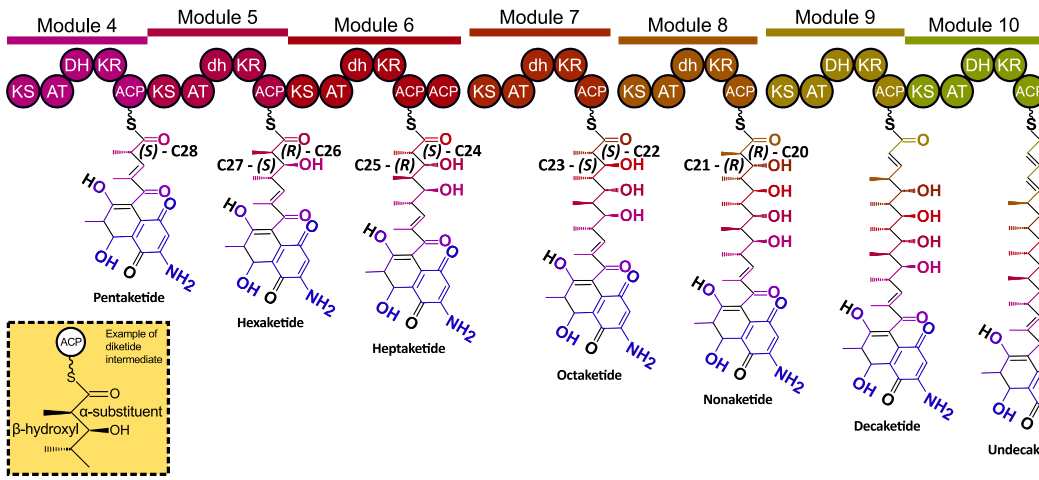
same module then the stereocontrol of the KR cannot be inferred (cryptic) from the structure of the molecule due to the formation of an unsaturation in the α,β-carbons of that intermediate. In most cases, the orientation of the α-substituent is maintained (Keatinge-Clay, 2012), since there are a few examples in which DH domains catalyse the isomerisation of the unsaturation, and the position of the double bond shifts from the typical α,β-carbons to the atypical β,γ-carbons (Kusebauch et al., 2010); furthermore, there is a report in which an atypical DH performs the epimerisation of the α-substituent (Keatinge-Clay, 2012).

The structure of prochaxamycin (which was deduced from the structure of chaxamycin A) was broken down into its intermediates and their chiral centres (α-carbon and β-carbon) were analysed in order to assess the type of stereocontrol exerted by the KR domains of the chaxamycin PKS (Figure 2.17). The configurations of the chiral centres were assigned by using CIP rules as in Keatinge-Clay, 2012 (Figure 2.17A, in B see column KR type). The stereocontrol applied by the KR domain of Module 1 on the diketide is likely lost during the formation of the naphthalenic ring and, therefore, is cryptic. The same occurs in the case of the KR domain of Module 4, since the DH domain is active and therefore it modifies the stereochemistry of the pentaketide. Cryptic stereocontrol of KR domains of Modules 9 and 10 was also observed, since they are accompanied by active DH domains.

The KR domain of Module 5 was classified as A1−type, since the configuration of β-C-27 and α-C-26 in the hexaketide is *S* and *R*, respectively. The KR domain of Module 6 is B2−type, since the configuration of β-C-25 is *R* and the α-C-24 is *S*. The KR domain of Module 7 is A2−type, since the configuration of the β-C-23 is *S* and the α-C-22 is *S*. The KR domain of Module 8 is B1−type, since the configuration of the β-C-21 is *R* and the α-C-20 is *R*.

Keatinge-Clay (Keatinge-Clay, 2007) defined a protocol to predict A− or B−type KR domains based on fingerprints present in their amino acid sequences (Figure 2.16B, examples are contained in a blue box). Alignment of the amino acid sequences of KR domains of Modules 0, 1, 3, 4, 5, 6, 7, 8, 9 and 10 of the chaxamycin PKS was performed to assess the predictability and usefulness of this method to predict stereocontrol of KR domains using the structure of chaxamycin A as example (Figure 2.17B). The results of the alignments indicate that the KR in the Loading Module is predicted as inactive since it lacks the main tyrosine residue (Figure 2.17B, Fingerprint 4) among others. The KR domain of Module 3 has the tyrosine residue mutated to serine (Figure 2.17B, Fingerprint 4) and

A



B

KR domain	LDD loop (100; 1)	Catalytic region (150, 160; 2, 3, 4, 5)	Fingerprint	Predicted KR type	KR type*
Ery2(A1)	HAAGLPQQVAI	SSGAGVWGSARQGAYAAGNA	**1.** No LDD; **2.** W; **3.** No H	A1	A1
Amp1(A2)	HTAAVIELAAL	SSTAGMWGSGVHAAYVAGNA	**1.** No LDD; **2.** W; **3.** H	A2	A2
Ave1(B1)	HTAGILDDATL	SSAAATFGAPGQANYAAANA	**1.** LDD; **5.** No P	B1	B1
Ery1(B2)	HAAATLDDGTV	SSFASAFGAPGLGGYAPGNA	**1.** LDD; **5.** P	B2	B2
CxmA_KR-0	HTADDPELALF	- - - - - - - - - - - - - - - - - - - -	-	-	-
CxmA_KR-1	HTAGVLDDGVL	SSAAGVLGNPGQGNYAAANA	**1.** LDD; **5.** No P	B1	N.d.[1]
CxmA_KR-3	HAADLTFTGTV	SSIAGIWGGGGQGPSAAANA	**4.** No Y	C1	C1
CxmB_KR-4	HLAGVLDDGVI	SSAAALMGSAGQGNYAAANA	**1.** LDD; **5.** No P	B1	N.d.[2]
CxmB_KR-5	HTAGVLDAGVV	SSVSGVFMAAGSGSYAAANA	**1.** No LDD; **2.** No W	A	A1[3]
CxmB_KR-6	HTAGVFDDGVI	SSVAGVVGGGGQGNYAAANA	**1.** FDD; **5.** No P	B1	B2[3]
CxmC_KR-7	HTAGVFDAGVI	SSASSLFLGAGSGGYAAANA	**1.** No LDD; **2.** No W	A	A2[3]
CxmD_KR-8	HTAGVLDDGVI	SSASGVFGSAGQGNYAAANA	**1.** LDD; **5.** No P	B1	B1[3]
CxmE_KR-9	HTAGVLDDGVI	SSASGVFGSAGQGNYAAANA	**1.** LDD; **5.** No P	B1	N.d.[4]
CxmE_KR-10	HTAGVADAGVI	SSVSGVFMGAGSGSYATANA	**1.** No LDD; **2.** No W	A	N.d.[4]

C

KR domain	LDD loop (97; 1)	Catalytic region (150; 2, 3, 4, 5)	Fingerprint	Predicted KR type	KR type*
Rif_KR7	HTAGVFDA	SSASSIFMGAGSGGYAAANA	**1.** No LDD; **2.** F; **3.** S	A	A2[5]
Rif_KR10	HTAGVPDA	SSVSAVFMGAGSGSYAAANA	**1.** No LDD; **2.** F; **3.** S	A	A2[5]
CxmB_KR-5	HTAGVLDA	SSVSGVFMAAGSGSYAAANA	**1.** No LDD; **2.** F; **3.** S	A	A1[3]
CxmC_KR-7	HTAGVFDA	SSASSLFLGAGSGGYAAANA	**1.** No LDD; **2.** F; **3.** S	A	N.d.[4]
CxmE_KR-10	HTAGVADA	SSVSGVFMGAGSGSYATANA	**1.** No LDD; **2.** F; **3.** S	A	N.d.[4]

Figure 2.17: Prediction of the stereocontrol exerted by KR domains on the chiral centres of chaxamycin polyketide intermediates based on amino acid fingerprints. (**A**) Chirality of the α-carbon and β-carbon of the polyketide intermediates in the proposed biosynthesis of chaxamycins. The chirality was assigned using CIP rules as in Keatinge-Clay, 2012. For simplicity, intermediates synthesised by Modules 4 to 10 are shown. One example of the positions of the α-carbon and β-carbon in a polyketide intermediate are shown in the yellow box. (**B**) Location of amino acid fingerprints. Alignment of the amino acid sequences of the LDD loop and the catalytic region of a KR domain were assessed to predict its stereocontrol. Bold numbers on the top of the alignment indicate the position of the fingerprints as in Keatinge-Clay, 2007. KR domains of known stereocontrol are contained in a blue box as an example: Ery2, KR of module 2 of erythromycin PKS is A1−type; Amp1, KR of module 1 of amphotericin PKS is A2−type; Ave1, KR of avermectin PKS is B1−type; Ery1. KR of module 1 of erythromycin PKS is B2−type. (**C**) Alignment of amino acid sequences of non-canonical A2−type KR domains, described for the rifamycin PKS (orange box), with KR domains of Modules 5, 7 and 10 of the chaxamycin PKS. *KR type assigned according to experimental evidence; [1]not possible to identify, stereochemistry is likely lost after the formation of the naphthalenic ring; [2]not possible to identify, stereochemistry modified by the DH domain in Module 4; [3]based on chemical structure of prochaxamycin; [4]not possible to identify, stereochemistry is modified by the DH of Modules 9 and 10 on the decaketide and undecaketide, respectively; [5]stereocontrol determined experimentally (You et al., 2013).

therefore both are predicted as a non-functional, C1−type KR. This is consistent with the known chemical structure of chaxamycin A.

The KR domain of Module 1 is thought to reduce the ketone group of the AHBA moiety in the diketide. Fingerprint analysis classifies it as B1−type, however, there is no evidence in the structure of chaxamycin A to confirm this prediction, since the stereochemistry of the chiral centres are probably lost during the formation of the naphthalenic ring by Cxm19. The KR domain of Module 8, on the other hand, is predicted as B1−type, since it possesses the LDD motif (Figure 2.17B, Fingerprint 1) and no proline residue in Fingerprint 5 (Figure 2.17B). This is consistent with the stereochemistry of the nonaketide.

The stereochemistry of the α-substituent and the β-hydroxyl in the heptaketide in Module 6 indicates that KR domain is B2−type, but the prediction based on fingerprints suggested B1−type. Although the LDD motif of the KR domain of Module 6 is not fully conserved, since it possesses the leucine residue mutated to phenylalanine, but the invariant second aspartate residue is present in this motif to classify it as B−type KR (Figure 2.16B, Fingerprint 1). This KR domain also lacks proline residue in Fingerprint 5 (Figure 2.17B) and therefore it was predicted as B1−type, according to this methodology.

According to the stereochemistry of the chiral centres of the hexaketide and octaketide, KR domains of Module 5 and 7 are A1− and A2−type, respectively. The absence of the second, and most important, aspartate residue in the LDD motif confirms that they could be classified as A−type, but no other Fingerprints were present in their sequences for further classifications.

In the case of the rifamycin PKS, the prediction of the stereocontrol based on fingerprints of Rif−KR7 and Rif−KR10 was not conclusive (You et al., 2013). In the case of Rif −KR7, the configuration of the chiral centres in the rifamycin intermediate (octaketide) indicates that the configuration of the α-carbon and β-carbon was *S* and *S*, respectively, classifying it as A2−type. However, the prediction of stereocontrol of Rif−KR7 based on fingerprints allows its classification as A−type, since it lacks the LDD motif (Figure 2.17C, Fingerprint 1), but no tryptophan (Fingerprint 2) nor histidine (Fingerprint 3) residues were found to conclude the final classification of this domain (Figure 2.17C). On the other hand, Rif−KR10 of the PKS was accompanied by an active DH, then its stereocontrol is cryptic. Rif−KR7 and Rif−KR10 were experimentally validated as A2−type (chirality of α-carbon was *S* and β-carbon was *S*), since they produced a (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-ACP (You et al., 2013). Alignment of the amino acid sequence of Rif −KR7 and Rif−KR10 with the KR domains of Modules 5, 7 and 10 of the chaxamycin PKS revealed that all of them did not have the diagnostic tryptophan residue in Fingerprint 2 nor histidine residue in Fingerprint 3 and instead, they had a phenylalanine and a serine residue, respectively (Figure 2.17C). In the case of the chaxamycin PKS, fingerprints found in the KR domain of Module 7 are identical to those found in the A2−type Rif−KR7, which makes plausible its classification as A2−type and also is consistent with the stereochemistry observed in the ocatketide intermediate. According to the structure of chaxamycin A, the stereocontrol of the KR domain of Modules 5 is A1−type and the KR

of Module 10 is cryptic. Fingerprints found in the amino acid sequences of KR domains of Modules 5 and 10 were not identical to those found in Rif−KR7 and Rif−KR10, therefore it is not conclusive whether or not they belong to the A2−type KR domains based on their amino acid sequences (Figure 2.17).

Thus the predictability of the rules currently available for assigning stereocontrol on these particular KR domains seems not infallible and needs further analysis.

2.5.3 Transcriptional regulation of the chaxamycin biosynthesis gene cluster

The chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 encodes for two putative transcriptional regulatory proteins, CxmY and CxmZ (Table 2.1). CxmY is a protein of 433 amino acids that only possesses a "helix-turn-helix" (HTH; Pfam accession: PF00196) domain in the C-terminus and no other domain. Most protein members that harbour HTH domains function as sequence-specific DNA binding domains, such as in the NarL/FixJ RRs (Galperin, 2010). BLASTp analysis using the amino acid sequence of CxmY revealed that it shares homology with Rif-Orf36 of *A. mediterranei* S699 (43%; AAS07758) and Orf5 of *Streptomyces* sp. CS, a transcriptional regulator located in the *nat* gene cluster (47%; NCBI accession: ADM46353), both possess the HTH domain at the C-terminus of the protein, but their function in the biosynthesis of rifamycins and naphthomycins have not been determined, and the role of CxmY in the chaxamycin biosynthesis needs to be elucidated.

CxmZ is another regulatory protein of 246 amino acids encoded in the chaxamycin biosynthesis gene cluster that harbours two domains: a "winged HTH" domain in the C-terminus (Trans_reg_C family; Pfam accession: PF00486), which is found in OmpR/PhoB RRs (Galperin, 2010), and a "receiver domain" in the N-terminus (Receiver_domain family; Pfam accession: PF00072), which is present in RRs that receives the signal from the sensor partner in a two-component systems. The "receiver domain" of CxmZ is a homologue to that present in histidine kinase two-component systems, however, the conserved aspartate residue present in typical RRs such as in PhoP of *S. coelicolor*, is mutated to serine in CxmZ (box I in Figure Appendix H.10). In addition, CxmZ does not possess the threonine residue that is also present in typical RRs (box III in Figure Appendix H.10), which implies that CxmZ is unlikely to be phosphorylated as typical RRs. Instead, it would belong to atypical RRs, since it does not contain all the conserved amino acid residues to be phosphorylated as a typical receiver domain (Liu et al., 2013). In addition, no sensory histidine-kinase protein was found encoded near to the chaxamycin biosynthesis gene cluster.

In principle, atypical RRs can be activated by binding to the end product or late biosynthetic intermediates of secondary metabolites, such as antibiotics (Liu et al., 2013). CxmZ shares homology with several atypical RRs that act as transcriptional activators in biosynthesis gene clusters, such as JadR1 of the jadomycin biosynthesis gene cluster of *S. venezuelae* (41%; Wang et al., 2009); Aur1P of the auricin biosynthesis gene cluster of *S. au-*

reofaciens (41%; Novakova et al., 2005); LanI of the landomycin biosynthesis gene cluster of *S. cyanogenus* and LndI of the landomycin E biosynthesis gene cluster of *S. globisporus* (39% and 38%, respectively; Rebets et al., 2008). Protein alignment of these sequences can be found in Figure Appendix H.10. The regulatory role of CxmZ in the biosynthesis of chaxamycins needs to be experimentally determined.

2.5.4 Immunity and export of chaxamycins in *S. leeuwenhoekii* C34 and *S. coelicolor*

2.5.4.1 Export of chaxamycins

S. leeuwenhoekii C34 and *S. coelicolor* strains have shown resistance to the presence of chaxamycins in bioassays experiments (Figures 2.11 and 2.12). In addition, since *S. leeuwenhoekii* C34 is able to excrete chaxamycins A−D into the supernatant, it implies the existence of an excretion system as occurs with the rifamycins in *A. mediterranei* S699 to avoid the accumulation of the antibiotic inside the cell. The rifamycin biosynthesis gene cluster encodes for a transporter RifP that belongs to the major facilitator superfamily (MFS) drug:H^+ antiporter, encoded in the *rifP* gene. Partial silencing of *rifP* using an antisense cassette resulted in a 70% decrease in the extracellular level of rifamycin B (Absalón et al., 2007). The expression of MFS drug antiporters is tightly regulated by transcriptional regulators (Kumar et al., 2013) and indeed, *rifP* is likely repressed by a TetR-like transcriptional regulator encoded by *rifQ*, which is immediately downstream to it.

The chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 and pIJ12853 do not possess a homologue to *rifP*. Homologues to RifP were searched for in the rest of the genome of *S. leeuwenhoekii* C34 by BLASTp analysis. Two homologue proteins to RifP were found, Sle_08130 (NCBI accession: CQR60276; 60% identical to RifP) and Sle_37020 (NCBI accession: CQR63162; 62% identical to RifP), which are located 249,321 nt to the right and 3,092,516 nt to the left of the chaxamycin biosynthesis gene cluster, respectively. Homologue to RifQ, the transcriptional regulator that accompanies RifP, was also searched for by BLASTp and Sle_37030 was found with the highest homology (43% identity; NCBI accession: CQR63163); interestingly, *sle_37030* was located immediately downstream to *sle_37020* (a homologue to *rifP*), suggesting that the system Sle_37020-Sle_37030 could be more related to the excretion system RifP-RifQ of *A. mediterranei* S699.

As *S. coelicolor* M1650 produces chaxamycins, since it carries pIJ12853, it should encode for a chaxamycin excretion system, also a homologue to RifP, elsewhere in its genome. One homologue of RifP was found in the *S. coelicolor* A3(2) genome, Sco4024 (62% identity; NCBI accession: CAC32364), which is a putative integral membrane efflux protein and, as well as Sle_37020 and Sle_08130, Sco4024 belongs to the MFS (Pfam accession: PF07690). *Sal. arenicola* CNS-205 also encodes for a protein containing a MFS domain, Sare_2597 (57% identical to RifP, 62% to Sle_08130, 58% to Sle_37020, 61% to Sco4024; NCBI accession: ABV98438) that also occurs outside the *sare* biosyn-

thesis gene cluster. MFS possess 14 transmembrane domains (Kumar et al., 2013) and indeed, 14 transmembrane domains were predicted from the secondary structure for all of these proteins (Figure Appendix H.11)

2.5.4.2 Immunity to chaxamycins

Chaxamycins and rifamycins are structurally related antibiotics. Rifamycin B is able to inhibit bacterial growth through strong binding to the β-subunit of the DNA-dependent RNA polymerase, encoded by the *rpoB* gene. Only one copy of this gene was found in the genome of *A. mediterranei*, which is located downstream to the right end of the rifamycin biosynthesis gene cluster (964 nt downstream to *rif-orf36* or 2,805 nt downstream to *rifJ*). This RpoB (Ames_0654) features five residues in the rif I region, known to confer rifampicin-resistance to it and to several organisms that harbour some of these five residues (Campbell et al., 2001): glutamine (Q432), threonine (T434), isoleucine (I437), aspartate (D438) and asparagine (N447) (Figure 2.18A, residues highlighted in red). The cloning of *ames_0654* gene into the rifampicin-sensitive *Mycobacterium smegmatis* conferred rifampicin-resistance to that microorganism. In addition, three point mutations were separately introduced in the *rpoB* gene of *Myco. tuberculosis*, which is rifampicin sensitive: D430Q (variant 1), S436D (variant 2), and S445N (variant 3). Each variant of the *rpoB* gene was sufficient to confer rifampicin resistance to *Myco. smegmatis* (Floss and Yu, 2005).

The genome of *S. leeuwenhoekii* C34 encodes for one RpoB, Sle_29840 (77% identical to Ames_0654), as well as *S. coelicolor* A3(2), Sco4654 (76% identical to Ames_0654) and *Salinispora arenicola* CNS-205, Sare_4325 (78% identical to Ames_0654). The RpoB of *S. coelicolor* A3(2) is rifampicin-sensitive, but less than most bacterial RNA polymerases (Chater, 1974). This RpoB possesses mutations in amino acid residues 432, 434, 437 and 438 (positions in Ames_0654) by aspartate, asparagine, leucine and serine, respectively, that would confer the rifampicin-sensitive phenotype (Figure 2.18A). These mutations were also found in the respective residues of the RpoB of *S. leeuwenhoekii* C34 and in the RpoB of the rifamycin-producer *Sal. arenicola* CNS-205, excluding threonine 434 (positions in Ames_0654) that is present in the later organism. The residue asparagine 447 (position in Ames_0654) was conserved across those RpoB, which alone has been found to confer resistance to *Myco. smegmatis* (Floss and Yu, 2005) (Figure 2.18A). The RpoB of *M. luteus* has all of the diagnostic residues (Figure 2.18A) mutated making this strain fully rifampicin-sensitive and also chaxamycin-sensitive, which is correlated with agar well diffusion bioassays (Figure 2.18B). In *S. leeuwenhoekii* C34, *rpoB* (nt 3,577,583 and 3,581,068) lies 1.29 Mb away from the chaxamycin gene cluster (nt 1,211,049 to 1,289,829), and its product does not have all the amino acid residues required for resistance to rifamycin at the expected positions (Figure 2.18), instead it possesses the same amino acids as *S. coelicolor* A3(2) RpoB, which is sensitive to rifamycin (Chater, 1974). Thus, it seems unlikely that chaxamycin may exert its antibiotic activity by binding to RpoB.

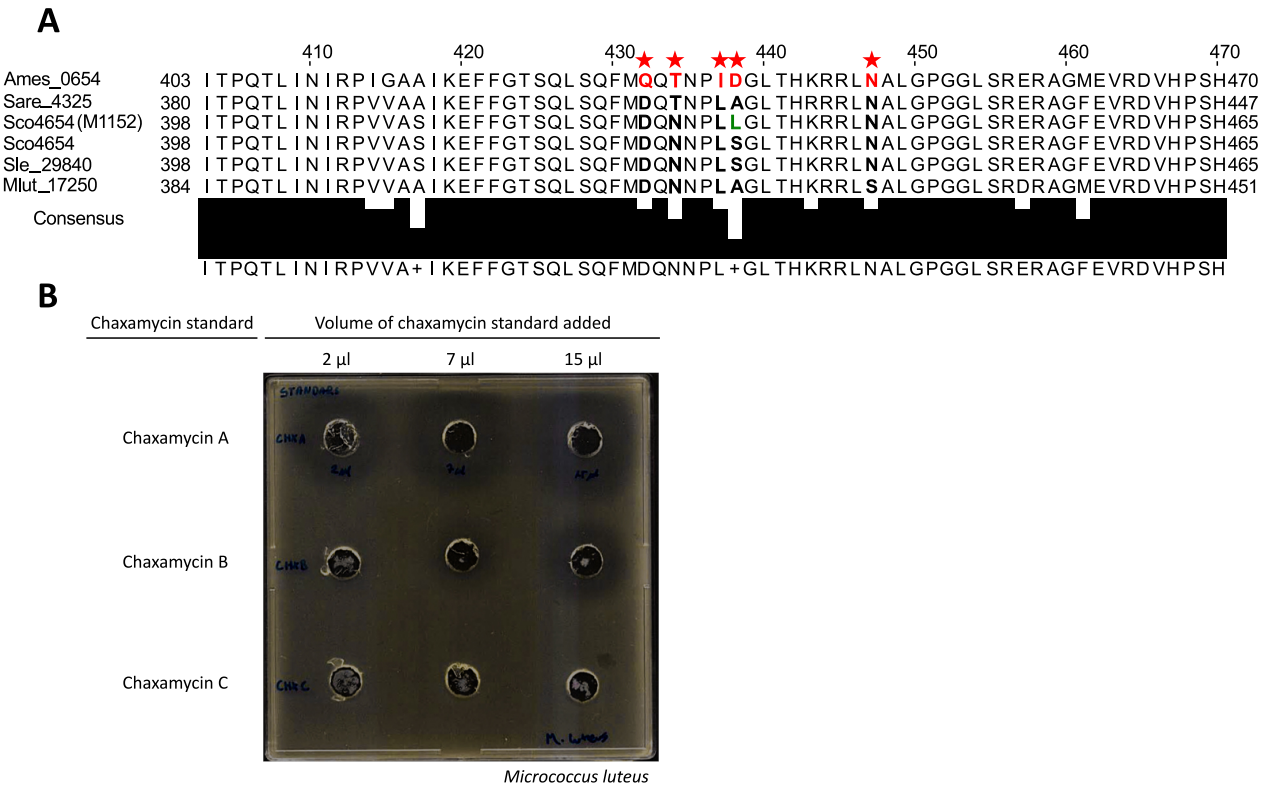


Figure 2.18: Alignment of amino acid sequences of bacterial RNA polymerase β-subunit (RpoB) to show the residues that confer resistance to rifampicin to *Amycolatopsis mediterranei* S699. (**A**) Five residues of *Amycolatopsis mediterranei* RpoB (Ames_0654), glutamine (Q432), threonine (T434), isoleucine (I437), aspartate (D438) and asparagine (N447) (highlighted in red) are present in the rif I region which confer resistance to rifampicin to several organisms (Campbell et al., 2001). *S. coelicolor* strain M1152 and the chaxamycin producer M1650 have the corresponding residue D438 mutated from serine to leucine, which confers rifampicin-resistance (residue in green) (Gomez-Escribano and Bibb, 2011). Sequences of RpoB were obtained from NCBI (accession): Ames_0654 of *A. mediterranei* S699 (Q9L637), Sco4654 of *S. coelicolor* A3(2) (CAB77428), Sare_4325 of *Salinispora arenicola* CNS-205 (ABW00107), Mlut_17250 of *Micrococcus luteus* (ACS31212) and Sle_29840 of *S. leeuwenhoekii* C34 (CQR62445); sequence of Sco4654 [S433L] was manually mutated, as in Gomez-Escribano and Bibb, 2011. (**B**) Agar well diffusion bioassays using *M. luteus* as indicator microorganism with chaxamycin A, B and C standards.

2.5.5 Genetic targets for metabolic engineering of *S. leeuwenhoekii* C34 towards increasing chaxamycin production

Chaxamycins are basically made up of AHBA, malonyl-CoA and (2*S*)-methylmalonyl-CoA precursor units. Its biosynthesis also incorporates a methyl group into the C-3 position of the naphthoquinone ring, using L-methionine as substrate. In particular, chaxamycin D possesses an *O*-acetylation at oxygen C-25, which probably comes from acetyl-CoA (or propionyl-CoA) (Xiong et al., 2005). To increase chaxamycin biosynthesis, two possible hosts for metabolic engineering could be considered. One is the native strain, which has shown to be amenable to genetic engineering, grows fast and is easy to work with. A second alternative is *S. coelicolor* M1650, a chaxamycin producer that has the entire gene cluster integrated into its genome; this strain has broader range of tools for genetic engineering and has a well-studied genetics. Targets for metabolic engineering will be discussed for both cases as follows.

2.5.5.1 Overview of the metabolic pathways involved in supplying the precursor molecules for chaxamycin biosynthesis

The biosynthesis of AHBA starts when UDP-α-D-glucose is converted, through several enzymatic steps, into D-kanosamine (Guo and Frost, 2002). The pool of UDP-α-D-glucose is generated from α-D-glucose-6-phosphate, an intermediate of glycolysis, by the enzyme Pgm that converts α-D-glucose-6-phosphate into α-D-glucose-1-phosphate, which is subsequently converted into UDP-α-D-glucose by the enzyme GalU. In addition, D-kanosamine biosynthesis requires availability of L-glutamine, which is generated by the enzyme GlnA that uses nitrogen sources from the intake of nitrate.

Other pathways are those related to the biosynthesis of malonyl-CoA and (2*S*)-methylmalonyl-CoA, the biosynthetic units needed for the assembly of the chaxamycin polyketide chain. Acetyl-CoA is generated from pyruvate, at the end of glycolysis, and represents a metabolic node that can divert carbon flux into malonyl-CoA, by the action of the enzyme complex Aac, and/or to citrate by the enzyme GltA. Therefore, carbon flux should be channelled from the acetyl-CoA node towards malonyl-CoA and away from citrate to promote the accumulation of malonyl-CoA intermediate in the cell, which can be achieved by over-expressing *acc* genes in *S. leeuwenhoekii* C34. (2*S*)-methylmalonyl-CoA could be synthesised by three pathways, including Pcc, Mct and Mcm-Mce pathways, being the later the most relevant source of (2*S*)-methylmalonyl-CoA in *A. mediterranei* S699 (Shao et al., 2015). In this Mcm-Mce pathway, Mcm converts succinyl-CoA into (2*R*)-methylmalonyl-CoA in a reversible reaction and Mce epimerises (2*R*)-methylmalonyl-CoA molecule into (2*S*)-methylmalonyl-CoA, which is the substrate recognised by PKSs.

2.5.5.2 Proposition of genetic modifications of *S. leeuwenhoekii* C34 oriented to increase chaxamycin production

To redirect the carbon flux from glycolysis into D-kanosamine biosynthesis, aiming at increasing the availability of UDP-α-D-glucose, the over-expression of the gene *sle_66460* that encodes for a Pgm is proposed. *pgm* gene from *S. coelicolor* (*sco7443*) was over-expressed in *S. argillaceus* to enhance production of the mithramycin antitumoral compound, but the increase in concentration of mithramycin was low (Zabala et al., 2013). They attributed this result to a decrease in the availability of malonyl-CoA precursors generated from glycolysis, which are needed for mithramycin biosynthesis. Zabala et al. (2013) proposed four genetic mutations to circumvent this problem: i) over-expression of the *S. coelicolor pgm*, ii) over-expression of the native *acc* gene, iii) inactivation of the glucose pyrophosphorylase gene, *glgCa*, which uses α-D-glucose-1-phosphate towards storage of carbohydrates, and iv) inactivation of the acyl-CoA:diacylglycerol acyltransferase gene *aftAa* that drains acetyl-CoA precursors to triacylglycerides. These mutations combined increased the concentration of mithramycin 4.7 times. Therefore, in addition to the over-expression of the *pgm* gene of *S. leeuwenhoekii* C34 (*sle_66460*), the deletion of the *S. leeuwenhoekii* C34 *glgCa* gene (*sle_07310*) is proposed. The over-expression of the genes that encode for the acetyl-CoA carboxylase subunits (*acc*) of *S. leeuwenhoekii* C34 (*sle_4760*, *sle_44630*

and *sle_27560*) could be other candidates that might lead to increased chaxamycin production. In fact, the over-expression of *acc* genes of *S. coelicolor* has led to a significant increase in actinorhodin production in that microorganism (Ryu et al., 2006).

Two strategies have been carried out to modify Mcm genes of the Mcm-Mce pathway, aiming at increasing the availability of (2*S*)-methylmalonyl-CoA, needed for the production of either erythromycin or rifamycin in *Sacc. erythraea* and *A. mediterranei* S699, respectively. Overproduction of erythromycin by *Sacc. erythraea* was achieved by duplicating Mcm genes, *mutA*, *mutB*, *meaB* and *mutR*, complemented with cultivation of *Sacc. erythraea* in an oil-based medium (Reeves et al., 2007). On the other hand, insertional disruption of the gene that encoded for the (2*S*)-methylmalonyl-CoA mutase large subunit (*mutB2* or *ram_39870*) in *A. mediterranei* S699 led to an unexpectedly high level of rifamycin B production, comparable to that obtained with the rifamycin B over-producer strain HP-130 (HP-130 is not a parental strain of S699) (Peano et al., 2014). To increase chaxamycin production in *S. leeuwenhoekii* C34, the disruption of *mutB* (*sle_13030*) could be considered as a first approach.

Apart from the genetic targets proposed, the modification of the composition of the culture medium is another strategy whose objective is to modify the metabolism of *S. leeuwenhoekii* C34 to improve chaxamycin production. Comparison of expression of genes involved in carbon and nitrogen metabolism in the rifamycin SV producer *A. mediterranei* U32 showed that the presence of 80 mM KNO_3 in the growth medium (glucose 1.0%, tryptone 0.2%, yeast extract 0.1%, beef extract 0.1%, glycerol 1.0%, w/v, pH 7.0) promoted biosynthesis of rifamycin SV and up-regulation of genes involved in its biosynthesis, such as those devoted to the production of UDP-α-D-glucose (*galU*), L-glutamine (*glnA*), AHBA (*rifL*, *rifK*, *rifMN*, *rifH*, *rifG* and *rifJ*), proansamycin X (*rifABCDEF*) and also those for nitrate uptake (*nasAC* and *nasBD*). In turn, genes that encode for enzymes that drain precursors like malonyl-CoA (fatty acid synthase, *fas*) or (2*S*)-methylmalonyl-CoA (methylmalonyl-CoA mutase, *mcm*), were down-regulated (Shao et al., 2015). Thus, one possibility is to modify the chaxamycin production medium, trying the addition of KNO_3 to evaluate differences in antibiotics production with the current chaxamycin production medium, modified ISP2.

Duplication of the chaxamycin biosynthesis gene cluster in the genome of *S. leeuwenhoekii* C34 is also a possible strategy since the entire cluster is cloned in pIJ12853 that possess the insertional site of ΦC31. Amplification of a biosynthesis gene cluster might be a common mechanism leading to increased antibiotic production in industrial strains (Yanai et al., 2006). In this process, the resistance genes are also duplicated when they are located within the biosynthesis gene cluster, which leads to enhance antibiotic resistance (Yanai et al., 2006; Olano et al., 2008). In the case of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34, no obvious chaxamycin resistance determinants were found within the chaxamycin biosynthesis gene cluster nor in pIJ12853, but possible transporters that might be involved in the self-resistance of *S. leeuwenhoekii* C34 to chaxamycins were found in its genome. The role of these transporters needs to be elucidated to duplicate them for an increased antibiotic resistance when increased levels of chaxamycins would be reached.

Strategies towards increasing chaxamycin production just in the native strain are oriented to the creation of the so-called "*super hosts*", by deletion of entire gene clusters that compete for precursors pools or by incorporating point mutations in the genome that have a pleiotropic effect on levels of secondary metabolite production (Gomez-Escribano and Bibb, 2011). One successful example was the development of a *S. coelicolor super host*, which begun with the deletion of four main endogenous gene clusters and then point mutations were introduced in its genome in both, RNA polymerase β-subunit, *rpoB* gene ([*C1298T*] that confers resistance to rifampicin) and ribosomal protein S12, *rpsL* gene ([*A262G*] that confers resistance to streptomycin). The heterologous production of chloramphenicol and congocidine revealed dramatic increases in antibiotic production compared with the parental strain. Similar mutations in homologous *S. leeuwenhoekii* C34 genes (*sle_29840* and *sle_29790*, respectively) could be introduced for a possible pleiotropic enhancement of chaxamycin production in *S. leeuwenhoekii* C34.

S. leeuwenhoekii C34 is known to produce desferrioxamine E, hygromycin A and chaxalactins (Rateb et al., 2011a,b), bioinformatic analysis has led to the identification of the locus of these putative biosynthesis pathways, which could be deleted by means of gene replacement. We propose the deletion of the locus of the desferrioxamine E biosynthesis cluster, which is between nt 5,237,176 to 5,244,356, spanning genes *sle_44550* to *sle_44600*. In addition, deletion of the hygromycin A biosynthesis gene cluster could also be carried out. This cluster spans a DNA region from nt 160,425 to 189,028 that includes genes *sle_01610* to *sle_01870*. The chaxalactin biosynthesis gene cluster is a third candidate to be deleted; this is located between nt 7,146,903 to 7,227,608 and comprises genes *sle_61390* to *sle_61480* (identification is described in Chapter 3). Although the biosynthesis of chaxalactins directly competes for the precursors pools with the chaxamycin biosynthesis gene cluster, chaxalactins are not detected in modified ISP2 medium and therefore its deletion would not make a significant impact.

Metabolic engineering strategies are stepwise and labour-intense processes that need a broad and careful evaluation of the targets to be modified in the microorganism. Computational tools like genome-scale metabolic models have been very successful in aiding the prediction of non-intuitive genetic targets that might lead to an increase in the production of chaxamycins. Currently, a genome-scale metabolic of *S. leeuwenhoekii* C34 that will include both primary and secondary metabolism is being developed (Razmilic et al., in preparation) and will become another tool to guide metabolic engineering strategies in *S. leeuwenhoekii* C34.

2.6 Conclusions

1. A genomic DNA region of 80.2 kb (nt 1,210,347 to 1,290,550), found by bioinformatic means, contains 27 genes that support chaxamycins A−D biosynthesis in *S. leeuwenhoekii* C34.

2. The domain architecture of the PKS subunits: CxmA, CxmB, CxmC, CxmD and CxmE, found in the 80.2 kb region, is consistent with the structure of chaxamycin A and, therefore, is related to chaxamycin biosynthesis.

3. Heterologous production of all chaxamycin species, achieved after cloning pIJ12853 into *S. coelicolor* M1152, confirms that the identified gene cluster is functional and does encode chaxamycin biosynthesis.

4. Feeding with exogenous AHBA to *S. leeuwenhoekii* M1653, a mutant of the AHBA synthase gene (Δ*cxmK*) that does not produce chaxamycins A−D, restores chaxamycin production, confirms a role for *cxmK* in AHBA synthesis and provides further evidence that the identified gene cluster encodes chaxamycin biosynthesis.

Chapter 3

Identification of the putative locus of the chaxalactin biosynthesis gene cluster in the genome of *Streptomyces leeuwenhoekii* C34

3.1 Abstract

A putative chaxalactin biosynthesis gene cluster has been identified in the genome of *S. leeuwenhoekii* C34 by bioinformatic means. Genome mining analysis led to the identification of a region of genomic DNA that spanned 80.7 kb (nt 7,146,903 to 7,227,608) and was comprised of 24 genes: five PKS subunits, named *cxlA*, *cxlB*, *cxlC*, *cxlD* and *cxlE*, whose predicted domains architecture matched perfectly with that needed to explain chaxalactins biosynthesis. From the proposed polyketide chain extension, chaxalactin A was postulated as the first product of this pathway, followed by chaxalactin B and chaxalactin C. A cytochrome P450 encoding-gene was found transcriptionally coupled to *cxlA*, which is thought to be part of the cluster and hence it was named *cxlF*. CxlF is an unusual cytochrome P450 that has mutated the absolutely conserved cysteine residue to alanine and has been hypothesised to perform the hydroxylation of chaxalactin A that gives rise to chaxalactin B probably aided by two Rieske domain-containing enzymes encoded in *sle_61490* and *sle_61500* genes. Two methyltransferases encoding-genes, *sle_61390* or *sle_61510* were predicted as *O*-methyltransferases and related to the *O*-methylation of chaxalactin B that gives rise to chaxalactin C. The participation of all these genes in the biosynthesis of chaxalactins remains to be experimentally determined.

3.2 Introduction

Chaxalactins A−C (Figure 3.1) are a rare class of 22-membered macrolactone polyketides synthesised by *Streptomyces leeuwenhoekii* C34. Chaxalactins display, through an unknown mechanism, selective antimicrobial activity against Gram-positive bacteria over Gram-negative bacteria (*Vibrio parahaemolyticus* and *Escherichia coli*). Chaxalactin A displayed high antimicrobial activity against *Bacillus subtilis* with MIC value of 0.2 µg/ml, chaxalactin B displayed high antimicrobial activity against *Listeria monocytogenes* and *Staphylococcus aureus* with MIC values of 5.0 µg/ml and 0.2 µg/ml, respectively, and chaxalactin C displayed high antimicrobial activity against *Staph. aureus* with MIC value of 0.8 µg/ml (Rateb et al., 2011b).

The group of 22-membered macrolactone polyketides comprises several compounds such as cytovaricin from *Streptomyces* sp. H-230 (Kihara et al., 1981), phthoramycin from *Streptomyces* sp. WK-1875 (Ōmura et al., 1988), compound A82548A and yokonolide from *S. diastatochromogenes* B59 (Hayashi et al., 2001), ushikulides A and B from *Streptomyces* sp. IUK-102 (Takahashi et al., 2005), atacamycins A−C from *S. leeuwenhoekii* C38 (Nachtigall et al., 2011), labilomycin (or pulvomycin) from *Streptoverticillium mobaraense* Tü 1063 (Wolf et al., 1978), etnangiene from the myxobacteria *Sorangium cellulosum* (Nett and König, 2007) and difficidin from *B. subtilis* (Zweerink and Edison, 1987). In particular, atacamycins are synthesised by a relative strain to the *S. leeuwenhoekii* clade, also isolated from the Atacama Desert (Alan Bull, personal communication).

The biosynthesis of chaxalactins is believed to be carried out by a type I PKS, different from that involved in the biosynthesis of chaxamycins (Rateb et al., 2011b). The chemical structure of chaxalactins indicates that the backbone of the molecule is composed of eight acetate and four propionate units, without using any specialised starting unit such as AHBA, needed for chaxamycin biosynthesis. Hence, different biosynthetic origins are proposed for these polyketides. The importance of chaxalactins lies in their strong and selective activity against Gram-positive bacteria but weak activity against Gram-negative bacteria (Rateb et al., 2011b). The aim of this work is the identification of the locus of the chaxalactin biosynthesis gene cluster in the genome of *S. leeuwenhoekii* C34, by bioinformatic means. This would allow the proposition of genetic modifications in the native strain, aiming at the experimental identification of the chaxalactin biosynthesis gene cluster.

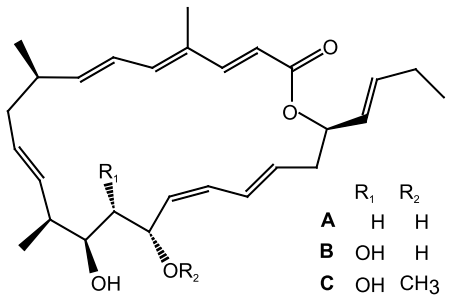


Figure 3.1: Chemical structures of chaxalactins A−C.

3.3 Materials and methods

3.3.1 Sequencing of the genome of *S. leeuwenhoekii* C34

High-quality genomic DNA was isolated from *S. leeuwenhoekii* C34 using standard procedures (Appendix F.1.3). Sequencing of the genome of *S. leeuwenhoekii* C34 was carried out using two technologies: Illumina MiSeq (Department of Biochemistry, University of Cambridge, Cambridge, UK) and PacBio RS II SMRT (The Genome Analysis Centre (TGAC), Norwich Research Park, Norwich, UK). The full process describing the assembling and manual curation process to generate a consensus sequence has been reported elsewhere (Gomez-Escribano et al., 2015). The consensus sequence of the genome was processed with Prodigal (Hyatt et al., 2010) to identify protein-coding sequences, which was automatically annotated with BASys (Van Domselaar et al., 2005) web server (https://www.basys.ca/). Analysis and curation of the consensus sequence was carried out with Artemis (Rutherford et al., 2000). All genome mining analysis were performed using this DNA sequence (accession number: LN831790).

3.3.2 Bioinformatic analysis

BLAST searches (Altschul et al., 1997) were performed at the NCBI server (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Local BLASTn, BLASTp or RPS-BLAST, against a database of genome/genes or proteins of *S. leeuwenhoekii* C34 was performed using NCBI package 2.2.26+ (Camacho et al., 2009) and SequenceServer (Priyam et al., in preparation) on a MacBook Pro OS X version 10.7.5. DoBISCUIT (Database of BIoSynthesis clusters CUrated and InTegrated; Ichikawa et al., 2013) was used to obtain curated and annotated versions of macrolide gene clusters. Conserved protein domains were searched for using the CDD Search at NCBI (Marchler-Bauer et al., 2015) or the InterPro web servers (Mitchell et al., 2015). Search for conserved motifs was aided by ScanPROSITE web server (Sigrist et al., 2012).

3.3.2.1 RPS-BLAST for finding PKS domains

Proteins of *S. leeuwenhoekii* C34 that encode for PKS domains were extensively searched for in its genome by using RPS-BLAST (Marchler-Bauer et al., 2015). A database of conserved protein domains was downloaded from the NCBI website (zipped file named "cdd.tar.gz" of 3.79 Gb, unzipped 27 Gb; ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd). This file contains Hidden Markov Models as position-specific scoring matrix for the extent of each domain consensus sequence.

A list of conserved protein domains present in PKS was obtained using the amino acid sequence of the subunits of the pikromycin PKS as model (DoBISCUIT accession: pikromycin). Each amino acid sequence of the pikromycin PKS was used as query in the CDD Search at NCBI website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?). The output of this search was a list with all the SMART domains (Letunic et al., 2015) present

in the pikromycin PKS (accession): AT (SM00827), ACP (SM00823), DH (SM00826), ER (SM00829), KR (SM00822), KS (SM00825) and TE (SM00824). A file with the list of these conserved domains was created and named “PKS_CDD.pn” (Appendix I.1). A custom RPS-database was created using these domains with the NCBI package 2.2.26+ (Camacho et al., 2009). The database was used to perform RPS-BLAST through all proteins encoded in the genome of *S. leeuwenhoekii* C34. A xml file with the results of the RPS-BLAST analysis was parsed using an algorithm developed in Python programming language version 2.7 (https://www.python.org/download/releases/2.7/) and the BioPython package version 1.65 (http://biopython.org/wiki/Download). The code developed is detailed in Appendix I.1. The result of the search was a list of PKS domain-containing proteins of *S. leeuwenhoekii* C34.

3.3.3 Generation of illustrations

Drawings of chemical structures were done in MarvinSketch version 6.2.0. ImageJ version 1.47 (http://imagej.nih.gov/ij) and Inkscape version 0.91 (https://inkscape.org) were used to create and edit drawings and illustrations.

3.4 Results

3.4.1 Bioinformatic identification of the chaxalactin biosynthesis gene cluster in the genome of *S. leeuwenhoekii* C34

Chaxalactins A–C (Figure 3.1) are 22-membered macrolactone polyketides produced by *S. leeuwenhoekii* C34. Chaxalactin A is made up by 12 building blocks, likely incorporated by a type I PKS composed by one Loading Module and 11 extensor Modules (Figure 3.2). The information derived from the structure of chaxalactin A will allow the identification of a putative chaxalactin PKS in the genome of *S. leeuwenhoekii* C34.

The Loading Module of the chaxalactin PKS would be in charge of incorporating the propionate starting unit, which is likely derived from (2*S*)-methylmalonyl-CoA (Dewick, 2009) (green bonds in Figure 3.2). Eight out of 11 Modules are predicted to incorporate acetate units derived from malonyl-CoA (orange bonds in Figure 3.2), and another 3 Modules are predicted to incorporate propionate units derived from (2*S*)-methylmalonyl-CoA (blue bonds in Figure 3.2). Interestingly, the structure of chaxalactin A reveals that the hydroxyl-group from the malonyl-CoA unit introduced by Module 7 in position C-9, was fully reduced (red arrow in Figure 3.2). This indicates that at least three *cis*-acting domains: KR, DH and ER should be present in Module 8 of the chaxalactin PKS. In addition, the macrolide ring observed in the structure of chaxalactin A is likely catalysed by a TE domain that is usually located at the C-terminus of the last Module of the PKS (Staunton and Weissman, 2001). Thus, a PKS composed by 12 modules was searched for in the genome of *S. leeuwenhoekii* C34 by RPS-BLAST analysis.

Five genes homologous to PKSs were found encoded in the genome of *S. leeuwenhoekii* C34: *sle_61410*, *sle_61430*, *sle_61440*, *sle_61450* and *sle_61460*. The protein domains predicted from the amino-acid sequences of these PKS-encoding genes showed an

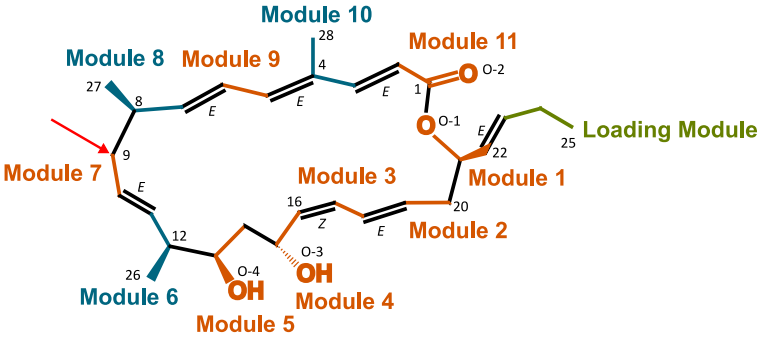


Figure 3.2: Biosynthesis origin of each extender unit present in chaxalactin A. Each extender unit is highlighted and accompanied by the module that adds that extender unit: (2*S*)-methylmalonyl-CoA starting unit (green); malonyl-CoA (orange); (2*S*)-methylmalonyl-CoA (blue). The numbers represent the position of the carbons and oxygen atoms in the molecule; letters *Z* and *E* indicate the configuration of the double bonds (Rateb et al., 2011b). The red arrow indicates that bond C-9 hydroxyl-group is completely reduced, suggesting the action of an enoylreductase domain in Module 8. Solid wedge indicates bond or group is projecting out towards the viewer; dashed wedge indicates bond or group is projecting away from the viewer; a line indicates that the bond is in the same plane.

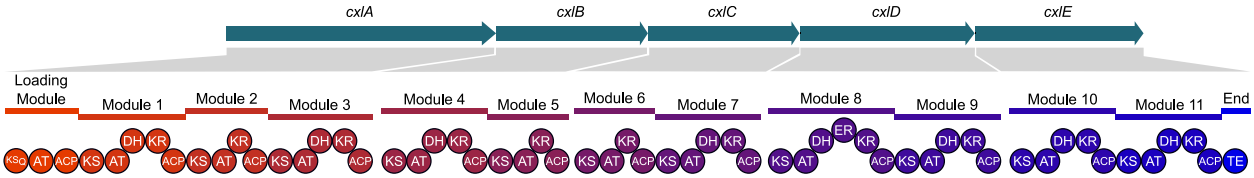


Figure 3.3: Domain architecture of the PKS found in the putative chaxalactin biosynthesis gene cluster. Conserved domains were predicted with the CDD Search at NCBI web server (Marchler-Bauer et al., 2015). KS_Q indicates that the domain has a conserved cysteine mutated to glutamine.

architecture identical to a type I PKS with one Loading Module, 11 extensor Modules and a TE domain in the C-terminus of the last PKS subunit, consistent with the modules needed to explain the biosynthesis of chaxalactin A (Figure 3.3). Thus, the PKS subunits encoded in the genes mentioned above were named (NCBI accession): CxlA (CQR65597), CxlB (CQR65598), CxlC (CQR65599), CxlD (CQR65600) and CxlE (CQR65601), respectively (Table 3.1). In addition, Module 8 present in CxlD included KR, DH and ER domains necessary for the full reduction of the hydroxyl group observed in position C-9, confirming that this PKS is highly likely to be in charge of the biosynthesis of chaxalactin A.

On the other hand, chaxalactins B and C are thought to be derived from chaxalactin A. The structure of chaxalactin B suggests that it is a product of a tailoring reaction that involves a hydroxylation at position C-14 of chaxalactin A. In turn, a C-13 *O*-methylation reaction at O-4 position of chaxalactin B would give rise to chaxalactin C. Therefore, it makes sense that genes encoding for a hydroxylase and a methyltransferase enzyme should be present near the chaxalactin PKS of *S. leeuwenhoekii* C34. Indeed, upstream to the putative chaxalactin PKS genes, the gene *sle_61400* encodes for a cytochrome P450 that seems to be transcriptionally coupled to *cxlA*. Cytochromes P450 are usually involved in hydroxylation reactions (Podust and Sherman, 2012) and therefore it is plausible that *sle_61400* is part of the cluster. Thus, the product of that gene was named CxlF (NCBI accession: CQR65596). Downstream to *cxlE* there are two genes (NCBI accession): *sle_61490* (CQR65604) and *sle_61500* (CQR65605) that encode for Rieske non-haem iron oxygenases, also involved in hydroxylation reactions (Que and Tolman, 2008), which could also be part of the cluster (Table 3.1).

In addition, two genes that encode for methyltransferases were found in the surroundings of the PKS genes (NCBI accession): *sle_61390* (CQR65595), located upstream to CxlF, and *sle_61510* (CQR65606), located downstream to the Rieske-containing proteins (Figure 3.4; Table 3.1). Three transport systems were found near the chaxalactin PKS: a drug/metabolite transporter permease (*sle_61320*; CQR65588) and two ABC

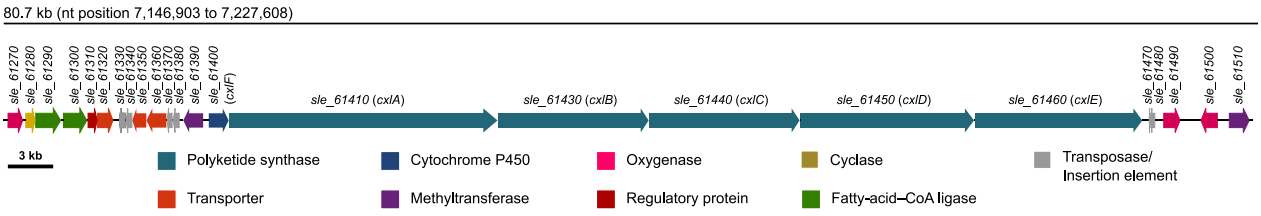


Figure 3.4: Proposed chaxalactin biosynthesis gene cluster of *S. leeuwenhoekii* C34.

transporters: *sle_61350* (CQR65591) and *sle_61360* (CQR65592). The putative transporter encoded in *sle_61320* was accompanied by a TetR regulatory protein (*sle_61310*; CQR65587), located 5,390 nt upstream to *cxlF*.

The limits of the chaxalactin biosynthesis gene cluster were assigned based on the presence of the genes that encoded for the necessary enzymes to explain the biosynthesis of chaxalactin A–C (Figure 3.4). Thus, *sle_61270*, that encodes for a nitronate monoxygenase, was assigned as the left end and *sle_61510*, that encodes for a methyltransferase, was established as right end of the cluster, spanning a DNA region of 80.7 kb (nt 7,146,903 to 7,227,608). Within this area, there are genes that encode for enzymes devoted to polyketide synthesis, tailoring reactions and transport.

Table 3.1: Proposed chaxalactin biosynthesis gene cluster and gene function.

Sle_ number	Cxl name[1]	Amino acids	Proposed function in chaxalactin biosynthesis	Homologue (% identity)[2]
Sle_61270		323	Nitronate monooxygenase	WP_030609607 *Streptomyces fulvoviolaceus* (89%)
Sle_61280		324	Metal-dependent hydrolase/cyclase	WP_030532061 *Prauserella rugosa* (77%)
Sle_61290		553	Acyl-CoA synthetase	WP_030609588 *Streptomyces fulvoviolaceus* (80%)
Sle_61300		519	Acyl-CoA synthetase	WP_011209630 *Nocardia farcinica* (69%)
Sle_61310		198	TetR transcriptional regulator	WP_020942979 *Streptomyces collinus* (86%)
Sle_61320		334	Drug/metabolite transporter permease	WP_020942978 *Streptomyces collinus* (82%)
Sle_61330		116	Transposase	WP_030700290 *Streptomyces griseus* (92%)
Sle_61340		136	Transposase	WP_030700291 *Streptomyces griseus* (94%)
Sle_61350		305	Sugar ABC transporter permease	WP_029150345 *Microbacterium indicum* (64%)
Sle_61360		308	Sugar ABC transporter permease	WP_029150344 *Microbacterium indicum* (56%)
Sle_61370		122	Transposase	WP_031156398 *Streptomyces xanthophaeus* (85%)
Sle_61380		128	Transposase	WP_031156395 *Streptomyces xanthophaeus* (86%)
Sle_61390		265	Tailoring: methyltransferase, C-13 *O*-methyltransferase (chaxalactin C)	ACY01395 *Streptomyces platensis* subsp. *rosaceus* (39%)
Sle_61400	CxlF	433	Tailoring: Cytochrome P450, C-14 hydroxylase (chaxalactin B)	WP_009947734 *Saccharopolyspora erythraea* (37%)
Sle_61410	CxlA	5,869	Polyketide synthase (Module0[3]: KS_Q-AT_{mmal}-ACP; M1: KS-AT_{mal}-DH-KR-ACP; M2: KS-AT_{mal}-KR-ACP; M3: KS-AT_{mal}-DH-KR-ACP)	WP_020871222 *Streptomyces rapamycinicus* (53%)

Sle_ number	Cxl name[1]	Amino acids	Proposed function in chaxalactin biosynthesis	Homologue (% identity)[2]
Sle_61430	CxlB	3,293	Polyketide synthase (M4: KS-AT_{mal}-DH-KR-ACP; M5: KS-AT_{mal}-KR-ACP)	WP_020871222 *Streptomyces rapamycinicus* (53%)
Sle_61440	CxlC	3,294	Polyketide synthase (M6: KS-AT_{mmal}-KR-ACP; M7: KS-AT_{mal}-DH-KR-ACP)	WP_020866322 *Streptomyces rapamycinicus* (54%)
Sle_61450	CxlD	3,788	Polyketide synthase (M8: KS-AT_{mmal}-DH-ER-KR-ACP; M9: KS-AT_{mal}-DH-KR-ACP)	WP_020871230 *Streptomyces rapamycinicus* (53%)
Sle_61460	CxlE	3,652	Polyketide synthase (M10: KS-AT_{mmal}-DH-KR-ACP; M11: KS-AT_{mal}-DH-KR-ACP; End: TE)	WP_020871231 *Streptomyces rapamycinicus* (50%)
Sle_61470		35	Transposase	WP_030661962 *Streptomyces cellulosae* (90%)
Sle_61480		103	Putative transposase	WP_012268487 *Streptomyces* sp. HK1 (43%)
Sle_61490		362	(2Fe-2S)-binding protein	WP_031521245 *Streptomyces* sp. NRRL F-5123 (42%)
Sle_61500		364	(2Fe-2S)-binding protein	WP_030231892 *Streptomyces* sp. NRRL S-350 (44%)
Sle_61510		277	Tailoring: methyltransferase, C-13 *O*-methyltransferase (chaxalactin C)	ACY01395 *Streptomyces platensis* subsp. *rosaceus* (51%)

[1] Abbreviation assigned to those proteins that are highly likely to be part of the chaxalactin biosynthesis gene cluster.
[2] BLASTp analysis using the non-redundant GenBank database.
[3] KS_Q indicates that the domain has a conserved cysteine mutated to glutamine.

3.4.2 Bioinformatic analysis of the PKS of the chaxalactin biosynthesis gene cluster

3.4.2.1 Analysis of a reading frame-shift in the chaxalactin PKS

Before the release of the final version of the genome of *S. leeuwenhoekii* C34, the bioinformatic search for the chaxalactin PKS was carried out using a draft version of the genome, which resulted in the identification of 6 putative PKS genes that encoded for PKS subunits (*sle_61410*, *sle_61420*, *sle_61430*, *sle_61440*, *sle_61450* and *sle_61460*). The predicted domain architecture of those PKS subunits were similar to that needed to explain chaxalactins biosynthesis. However, the PKS domains predicted from Sle_61410 and Sle_61420 had an unusual architecture, specifically a broken KS domain was found at the C-terminus of Sle_61410 and at the N-terminus of Sle_61420. This observation was attributed to an unexpected reading frame-shift in the genome sequence of *S. leeuwenhoekii* C34 (Figure Appendix I.2A).

Alignment of all Illumina reads on the genome sequence of *S. leeuwenhoekii* C34 revealed that the nucleotide sequence GTGGA had an extra guanine. This was carried out using software: Burrows-Wheeler Aligner, Smith-Waterman (BWA-SW; Li and Durbin, 2010) for the alignment and SAMtools (Li et al., 2009) for the single nucleotide poly-

morphisms identification (Juan Pablo Gomez-Escribano, personal communication). Thus, the GTGGA sequence was corrected to GTGA (Figure Appendix I.2B), fixing the reading frame-shift. Then, the formerly predicted protein sequences, Sle_61410 (2,887 amino acids) and Sle_61420 (2,953 amino acids) are part of a new protein, Sle_61410 (5,869 amino acids). This new Sle_61410 protein is part of the current version of the genome of *S. leeuwenhoekii* C34 and the predicted PKS domain architecture from that protein was as expected, without broken domains in it.

3.4.2.2 Sequence-based assessment of the functionality of the domains present in the chaxalactin PKS

The amino acid sequences of the domains present in each module of the putative chaxalactin PKS were analysed to predict their functionality. Active ACP domains possess a serine residue that harbours the 4'-phosphopantetheine prosthetic group attachment site, needed to shuttle extender units and polyketide intermediates from one module to the one downstream. All 12 ACP domains of the putative chaxalactin PKS possess that serine residue (Figure Appendix I.3).

AT domains are capable of recognising the appropriate extender unit that will be linked to the growing polyketide chain. The chaxalactin PKS possesses one AT domain in each of the 12 Modules. Eight AT domains were predicted to recognise malonyl-CoA as substrate since they contained the motifs GH**S**(I/V)G and HAF**H**, while 4 AT domains were predicted to recognise (2*S*)-methylmalonyl-CoA, including that of the Loading Module, based on the presence of the motifs GH**S**QG and YAS**H** (Table 3.2). All these domains were also predicted as active AT because of the presence of the catalytic residues serine and histidine in the first and second motif, respectively (Keatinge-Clay, 2012).

Table 3.2: Prediction of substrate specificity of AT domains in the chaxalactin PKS.

AT of module	Signature I[1]	Signature II[1]	Predicted substrate specificity	Extender unit incorporated in chaxalactin A molecule
0	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
1	GH**S**LG	HAF**H**	Malonyl-CoA	Malonyl-CoA
2	GH**S**LG	HAF**H**	Malonyl-CoA	Malonyl-CoA
3	GH**S**LG	HAF**H**	Malonyl-CoA	Malonyl-CoA
4	GH**S**LG	HAF**H**	Malonyl-CoA	Malonyl-CoA
5	GH**S**VG	HAF**H**	Malonyl-CoA	Malonyl-CoA
6	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
7	GH**S**VG	HAF**H**	Malonyl-CoA	Malonyl-CoA
8	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
9	GH**S**IG	HAF**H**	Malonyl-CoA	Malonyl-CoA
10	GH**S**QG	YAS**H**	(2*S*)-Methylmalonyl-CoA	(2*S*)-Methylmalonyl-CoA
11	GH**S**LG	HAF**H**	Malonyl-CoA	Malonyl-CoA

If signature I and II are GH**S**(I/V)G and HAF**H**, then AT recognises malonyl-CoA as substrate; if Signature I and II are GH**S**QG and YAS**H**, then AT recognises (2*S*)-methylmalonyl-CoA as substrate. Catalytic residues are in bold.

[1] Based on Keatinge-Clay, 2012 (Keatinge-Clay, 2012).

The linkage of the new extensor unit into the growing polyketide chain is catalysed by KS domains by means of Claisen condensation (Dewick, 2009). Elongating KS domains possess three residues that form the catalytic triad: a cysteine in the TA**C**SSS motif, a histidine in the EA**H**GTG motif and another histidine in the KSNIG**H**T motif (Keatinge-Clay, 2012; Fernández-Moreno et al., 1992). The chaxalactin PKS possesses 12 KS domains in total, of those, 11 are located from Module 1 to 11 and are predicted as elongating KS domains because of the presence of the cysteine residue in the catalytic triad (Figure Appendix I.4). The KS domain of the Loading Module has the reactive cysteine residue mutated to glutamine in the TA**C**SSS motif, which is a common mutation found in KS domains of the loading modules of PKSs. These KS domains are known as KS_Q domains.

Eleven KR domains were found in the chaxalactin PKS, distributed in Modules 1 to 11 (only the Loading Module lacks KR domain). KR domains are involved in the reduction of the β-keto group of the polyketide intermediate using NADPH, leaving a β-hydroxyl group as a result (Keatinge-Clay, 2012, 2007). Reductase-active KR domains can be inferred by the presence of three diagnostic regions that should be present for binding NADPH: a GxGxxG motif, a Swap linker and a WGxW motif (Garg et al., 2014). Also the presence of the catalytic triad lysine, serine and tyrosine should also be present in active KR domains (Reid et al., 2003). All the KR domains found in the chaxalactin PKS possess all the described motifs to classify them as reductase-active, which is consistent with the structures of chaxalactins A−C (Figure Appendix I.5).

After the reduction of the β-keto group of the polyketide intermediate by a KR domain, DH domains catalyse the dehydration of β-hydroxyl group of the same polyketide intermediate, leaving an unsaturation in the α,β position. Active DH domains possess in their amino acid sequences the conserved motifs **H**xxxGxxxxP and **D**(A/V)(V/A)(A/L)(Q/H), whose respective histidine and aspartate residues are part of the active site (Valenzano et al., 2010). The structure of chaxalactin A contains unsaturations at positions C-23−C-22 (*E*), C-19−C-18 (*E*), C-17−C-16 (*Z*), C-11−C-10 (*E*), C-7−C-6 (*E*), C-5−C-4 (*E*), C-3−C-2 (*E*), and a full reduction at C-9−C-8. This evidence indicates that dehydration reactions occurred on the polyketide intermediates catalysed by active DH domains of Modules 1, 3, 4, 7, 8, 9, 10 and 11 (Figure 3.2). The amino acid sequence of all these DH domains show the presence of the catalytic dyad (Figure Appendix I.6), hence they are predicted as active domains, which is consistent with the structure of chaxalactin A.

Particularly, Module 8 possesses three *cis*-acting modifying domains: KR, DH and ER. An ER domain is capable of reducing the α,β double bond that results from the reduction and dehydration of the β-ketone group by a KR domain and a DH domain, respectively. ER domains use NADPH as cofactor and are able to modify the stereochemistry of the α-substituent. A conserved NADPH-binding motif, HxAx(G/T)GV(G/A)(M/S)A is present in competent ER domains (Keatinge-Clay, 2012). After the enoyl-reduction occurs, the configuration of the α-carbon of the polyketide intermediate is modified to either, (2*S*)-configuration or (2*R*)-configuration. The presence of a conserved tyrosine residue indicates that the ER domain leaves the α-carbon of the polyketide intermediate in (2*S*)-configuration (the α-substituent in L-orientation) (Keatinge-Clay, 2012; Kwan et al., 2008). The amino acid sequence of the ER domain present in Module 8 possesses

the NADPH-binding motif and lacks the conserved tyrosine residue (Figure Appendix I.7), suggesting that this domain is ER-competent and is likely to leave the α-carbon of the polyketide intermediate in (2*R*)-configuration.

3.5 Discussion

The putative chaxalactin biosynthesis gene cluster was identified in the genome of *S. leeuwenhoekii* C34 by bioinformatic means. This cluster spans a region of 80.7 kb (nt 7,146,903 to 7,227,608) and is comprised of 24 genes whose hypothetical function will be described based on the presence of conserved protein domains presented in Table Appendix I.1.

3.5.1 Function of the genes present in the chaxalactin biosynthesis gene cluster of *S. leeuwenhoekii* C34

The core of the proposed chaxalactin biosynthesis gene cluster is composed of 5 genes: *cxlA*, *cxlB*, *cxlC*, *cxlD* and *cxlE* that encode for PKS subunits involved in the elongation of the chaxalactin polyketide chain.

3.5.1.1 Chaxalactin A is the unique product of the extension of the chaxalactin polyketide chain by the PKS

The Loading Module of the chaxalactin PKS is comprised of domains KS_Q, AT and ACP. A KS_Q domain is unable to catalyse the formation of a C-C bond as a normal elongating-KS domain would do, instead, they are able to initiate the elongation of the polyketide chain by catalysing the decarboxylation of acyl-CoA starting molecules (Long et al., 2002; Keatinge-Clay, 2012). For example, the KS_Q of the monensin PKS catalyses the strict decarboxylation of malonyl-CoA starting molecule, generating an acetyl-ACP required to continue the elongation of the polyketide chain (Bisang et al., 1999); the KS_Q of the pikromycin PKS catalyses the decarboxylation of a methylmalonyl-CoA molecule, generating a propionyl-ACP required to continue the elongation process (Witkowski et al., 1999; Kim et al., 2004). In the case of the chaxalactin biosynthesis, the extension of the polyketide chain would begin when the AT domain recognises the (2*S*)-methylmalonyl-CoA starting molecule that is subsequently decarboxylated by the KS_Q domain, giving rise to the propionate moiety observed in the structure of chaxalactin A. This molecule is passed on to Module 1, in which a malonyl-CoA unit is incorporated into the growing polyketide chain, forming a diketide intermediate. The β-ketone of this diketide is sequentially reduced and dehydrated by the KR and DH domains of Module 1, leaving an unsaturation in the α,β-carbon position (C22-C23 of chaxalactin A structure; Figure 3.2).

A malonyl-CoA unit is linked to the diketide by Module 2, forming a triketide intermediate. The KR domain of Module 2 reduces the β-ketone, leaving a β-hydroxyl group. Modules 3 and 4 have similar domain organisation: KS, AT, DH, KR and ACP, both AT domains recognise malonyl-CoA units that are incorporated into the growing polyketide chain to form tetraketide and pentaketide intermediates, respectively. These intermediates also undergo the reduction and dehydration of their respective β-ketone, giving rise to α,β-carbon unsaturations. The pentaketide is passed on to Module 5 that incorporates a malonyl-CoA unit that gives rise to a hexaketide. This Module harbours a KR domain that is in charge of

reducing the β-ketone of the hexaketide. Then, Module 6 links a (2*S*)-methylmalonyl-CoA extender unit to the growing polyketide chain to form a heptaketide, and its β-ketone group is reduced by the KR domain of Module 6.

Module 7 incorporates a malonyl-CoA to originate an octaketide, whose β-ketone is reduced and dehydrated by KR and DH domains, respectively. The AT domain of Module 8 incorporates a (2*S*)-methylmalonyl-CoA extender unit to produce a nonaketide. In addition to the KR and DH modifying domains of Module 8, this Module possesses an ER domain that is able to further reduce the double bond produced after the reduction and dehydration of the β-ketone of the nonaketide. ER is also able to modify the stereochemistry of the polyketide intermediate, which in all previous steps was governed by KR and DH domains.

Modules 9, 10 and 11 possess the same domain architecture: KS, AT, DH, KR and ACP. Both AT domains of Modules 9 and 11 recognise malonyl-CoA while the AT of Module 10 recognises the (2*S*)-methylmalonyl-CoA extender unit. The incorporation of the respective extender units by Modules 9, 10 and 11 give rise to decaketide, undecaketide and dodecaketide intermediates, respectively. Then, these three intermediates undergo the action of KR and DH domains, which leave an unsaturation in their respective α,β-carbon position. In the particular case of the undecaketide, the α-methyl group looses its stereochemistry because of the dehydration step carried out by the DH domain, which can be observed at the C-4 position in chaxalactin A (Figure 3.2).

The final step for the release and cyclisation of the chaxalactin dodecaketide intermediate implicates the catalytic action of a TE domain, which in the chaxalactin PKS is fused to the C-terminus of Module 11. The mechanism by which this occurs involves a catalytic triad composed of histidine, aspartate and serine residues (Tsai et al., 2001), which are present in the chaxalactin TE domain (Figure Appendix I.8). The aspartate residue stabilises the histidine residue, which acts as the catalytic base that accepts the proton from the serine residue, which attacks the dodecaketide-ACP substrate located on the last Module of the PKS, yielding a dodecaketide-*O*-serine intermediate on the TE domain. The NH of the residue next to the serine, may serve as the oxyanion stabilising residue. The proposed mechanism of cyclisation would involve the attack of the C-21 hydroxyl group, a secondary alcohol, to the C-1 of the dodecaketide intermediate, resulting the formation of a cyclic intramolecular ester lactone bond that gives rise to chaxalactin A.

Thus, chaxalactin A is postulated as the first molecule synthesised of the group of the chaxalactins.

3.5.1.2 Tailoring reactions are involved in the biosynthesis of chaxalactin B and chaxalactin C

Conversely to chaxalactin A that is a direct product of the polyketide chain extension, chaxalactin B and C are likely produced by tailoring or postsynthetic modifications made on the structure of chaxalactin A: chaxalactin B is a 14-hydroxychaxalactin A and chax-

alactin C is a 13-*O*-methylchaxalactin B, therefore chaxalactin B and chaxalactin C are hypothesised as the second and third products of this biosynthesis route, respectively.

Postsynthetic modifications are likely to be carried out by, at least, the product of the gene *cxlF* (*sle_61400*), a cytochrome P450 and either (or both) *sle_61390* or *sle_61510*, predicted as *O*-methyltransferases, part of the proposed chaxalactin biosynthesis gene cluster. CxlF encodes for a cytochrome P450 monoxygenase, harbouring a CypX domain (CDD accession: COG2124). Cytochromes P450 perform a variety of biological functions such as, hydroxylation, epoxidation, decarboxylation, furan ring formation, among others that generate diverse oxygenation patterns in secondary metabolites, with a high degree of regio-selectivity and stereo-selectivity for the substrate they use (Podust and Sherman, 2012).

In essence, members of the family of cytochromes P450 are haem-containing proteins, which means that coordinate a haem prosthetic group through a cysteine residue. A haem prosthetic group consists of a ferrous iron atom (Fe^{2+}) in the centre of a porphyrin ring that uses O_2 for its activation (Que and Tolman, 2008). In addition to haem, bacterial cytochromes P450 also require auxiliary redox partner proteins such as, small redox [2Fe-2S] iron-sulphur ferredoxin and a FAD or FMN-containing reductase to shuttle the electron to the cytochrome P450 haem iron and thus catalyse its biological function (Lamb and Waterman, 2013). Cytochromes P450 have several conserved residue that make up the haem-binding pocket: WxxxR in the C-terminus, the glutamate-asparagine-asparagine triad present in **E**xx**R** and PE**R**F motifs and a conserved cysteine residue in Gxxx**C**xG; also, a conserved threonine residue in A/G-Gxx**T** that is related to oxygen binding (Sezutsu et al., 2013). The cysteine residue in Gxxx**C**xG motif is the only absolutely conserved residue across all 15,000 cytochrome P450 sequences, and its role is to serve as an axial ligand for the iron residue in the haem-binding region, along with the glutamate-asparagine-asparagine triad (Lamb and Waterman, 2013).

The amino acid sequence of CxlF suggests that it is not a typical cytochrome P450 (Figure Appendix I.9). CxlF was scanned for the conserved haem-binding motif in Scan-PROSITE web server (Sigrist et al., 2012), but the result indicated that the motif was absent. Apparently, CxlF lacks the entire conserved GxxxCxG motif (Figure Appendix I.9A, motif V). Both, ExxR and PERF motifs were present, but in the case of the A/G-GxxT motif, only the second glycine residue was conserved. Mutations in these positions of the conserved A/G-GxxT motif have been observed in other cytochromes P450: the A/G position showed a mutation to threonine, which has also been observed in the CalO2 enzyme (Uniprot accession: Q8KND6), a putative orsellenic acid P450 oxidase involved in calicheamicin biosynthesis by *Micromonospora echinospora* (McCoy et al., 2009). The position of the threonine residue in the A/G-GxxT motif shows a mutation to asparagine that has also been observed in the OxyB enzyme (Uniprot accession: Q8RN04), related to the coupling of aromatic side chains in vancomycin biosynthesis by *A. orientalis* (Zerbe et al., 2002).

Due to the increase in the number of sequenced genomes, several putative cytochrome

P450-encoding enzymes that lack the conserved cysteine have been found, suggesting that there could be potential exceptions to the rule. For example, in a putative cytochrome P450 of *Trichoderma reesei* QM6a (NCBI accession: EGR45174) the cysteine residue has been mutated to asparagine, while in a hypothetical protein of *Magnaporthe oryzae* 70-15 (NCBI accession: EHA56235), the cysteine residue has been mutated to phenylalanine (Chen et al., 2014). Whether these atypical cytochromes P450 are active is unknown. Alignment of the amino acid sequences of those atypical cytochromes P450 with CxlF indicates that an alanine residue could be in the position of the cysteine in the GxxxCxG motif (Figure Appendix I.9B). In any case, the participation of CxlF in the hydroxylation of chaxalactin A to give rise to chaxalactin B remains to be determined by in-frame mutational analysis.

Sle_61490 and Sle_61500 are two enzymes encoded in genes located 1,372 nt and 3,811 nt downstream of *cxlE*, respectively. Each of these enzymes contain a Rieske domain in the N-terminus and therefore are likely to be oxygenases that incorporate hydroxyl group into carbon-hydrogen bonds, using molecular oxygen as substrate and NAD(P)H as the reductant (Que and Tolman, 2008). Rieske-containing proteins lack the haem prosthetic group present in cytochromes P450, but instead they possess a non-haem iron atom that mediates the oxygenation reaction and also possess an iron−sulphur, [2Fe−2S], cluster that allows the storage of electrons (Ferraro et al., 2005).

Conserved motifs present in Rieske domain-containing proteins include the **C**x**H**x_{17}-**C**xx**H** that binds the [2Fe-2S] cluster and the **D**/E-x**H**xxxx**H** motif that coordinates a non-haem iron (Sydor et al., 2011). Both motifs are present in Sle_61490 and Sle_61500, indicating that they are likely active oxygenases (Figure Appendix I.10). Rieske domain-containing enzymes perform the incorporation of one or two hydroxyl moieties in reactions such as, dihydroxylation of double bonds in polyketides (Liu et al., 2015), oxidative cyclisation of natural products (Sydor et al., 2011) or monooxygenation of cholesterol (Capyk et al., 2009). One example of the action of Rieske domain-containing enzymes comes from the production of galbonolides A and B polyketides by *S. galbus*. The galbonolide biosynthesis gene cluster (*galABCDE*) harbours a cytochrome P450 and a N-terminus Rieske [2Fe−2S] domain-containing protein, encoded in *galD* and *galE*, respectively. Deletion of *galD* resulted in the abolition of galbonolide A and B production, and *in trans* complementation with *galABCDE* restored the production of these compounds. The authors suggested that the dihydroxylation observed in the structures of galbonolide A and B is likely incorporated by the GalDE system (Kim et al., 2014). The participation of Sle_61490 and Sle_61500 in aiding the hydroxylation step of chaxalactin A has to be experimentally verified.

Four genes, *sle_61270*, *sle_61280*, *sle_61290* and *sle_61300* are located on the left end of the chaxalactin biosynthesis gene cluster and encode for a nitronate monooxygenase, a cyclase and two acyl-CoA synthetases, respectively. However, no obvious function in the chaxalactin biosynthesis has been attributed for these enzymes. For example, Sle_61270 harbours the 2-nitropropane dioxygenase domain that is also present in a homologous of *Pseudomonas aeruginosa* and with which shares 68% identity. The 2-nitropropane dioxygenase of *P. aeruginosa*, as well as others that contain the same protein

domain, is devoted to catalyse the oxidative denitrification of nitroalkanes, widely used in industry, to their corresponding carbonyl compounds and nitrites (Ha et al., 2006). Alignment of the amino acid sequences of these 2-nitropropane dioxygenases with Sle_61270 (Figure Appendix I.11), showed that Sle_61270 has the catalytic histidine and all conserved residues for interacting with FMN, implying that this enzyme has a function different from that needed to explain chaxalactin biosynthesis and, therefore, this could not be part of the cluster. In any case, the participation of the enzymes encoded in the four genes mentioned in the biosynthesis of chaxalactins remains to be shown.

Sle_61390 and Sle_61510 are enzymes of 265 and 277 amino acids, respectively, that harbour an AdoMet_MTases domain (Conserved Domain accession: CD02440). According to the InterPro database (Mitchell et al., 2015), both proteins harbour a PKS_MeTfrase domain, usually found in PKSs (InterPro accession: IPR020803). The presence of the AdoMet_MTases domain in Sle_61390 and Sle_61510 suggests that they are SAM-dependent MTs. Three conserved motifs that are present in functional MTs were found in both Sle_61390 and Sle_61510: motif I is (L/I/V)(V/L)(E/D)(V/I)**G**(C/G)**G**(P/T)**G** that contains the GxGxG sequence that is the nucleotide-binding site for the SAM substrate, motif II is (G/P)(T/Q)(A/Y/F)**D**A(Y/V/I)(I/F)(L/V/C) in which the central aspartate residue is invariant and forms hydrogen bonds with both hydroxyls of the SAM ribose and motif III is LL(K/R)P**GG**(L/I/R)(I/L)(V/I/F/L)(L/I) whose central glycines are highly conserved and at least one of them is present across all MTs (Schubert et al., 2003; Kagan and Clarke, 1994) (Figure Appendix I.12).

SAM-dependent methyltransferases can be classified according to the atom at which they deliver the methyl group (Schubert et al., 2003). Thus, there exist *N*-MTs, *C*-MTs or *O*-MTs if they methylate a nitrogen, carbon or oxygen atom of a molecule, respectively. The presence of certain residues in the amino acid sequence of PKS-MTs allows their classification (Ansari et al., 2008). Ansari et al. (2008) propose that the presence of the second glycine residue in motif I and the aspartate residue in motif II suggest that they are likely *O*-MTs. Both Sle_61390 and Sle_61510, seem to fit these requirements and it is proposed that both are *O*-MTs that could methylate chaxalactin B to give rise to chaxalactin C (Figure Appendix I.12). The participation of either Sle_61390 or Sle_61510 in the biosynthesis of chaxalactins needs to be determined by mutational analysis in *S. leeuwenhoekii* C34.

3.5.2 Genes that might encode for transporters involved in the export of chaxalactins in *S. leeuwenhoekii* C34

The putative chaxalactin biosynthesis gene cluster harbours three genes related to transport, located near the chaxalactin PKS (Table 3.1). Sle_61320 encodes for a DMT permease, Sle_61350 and Sle_61360 encode for sugar ABC transporter permeases.

Sle_61320 contains 2 EamA protein domains (Pfam accession: PF00892) with 10 transmembrane domains (10−TMS). According to Jack et al. (2001) Sle_61320 could be

classified into the 10−TMS DME family, which is one member of the 14 families of transporters of the DTM superfamily. EamA domain, formerly known as domain of unknown function 6 was named after the *O*-acetyl-serine/cysteine export gene in *E. coli* and some EamA domain-containing proteins have been characterized as nucleotide-sugar transporters (Västermark et al., 2011). Recently an EamA domain-containing protein of 8 transmembrane domains was experimentally verified as a riboflavin transporter in *Rhizobium leguminosarum* (García Angulo et al., 2013). The search for EamA domain-containing proteins in the UniProt database and filtered by "Swiss-Prot Reviewed" showed that almost all enzymes retrieved from the search were related to sugar or amino acid transport, none related to drug transport (short link: http://goo.gl/6P9SkR). In addition, the search for an EamA domain-containing protein in all transporters located within a biosynthetic gene cluster deposited in DoBISCUIT database (sequences sent by Ms. Akane Kimura from DoBISCUIT) resulted in no-hit for this domain, suggesting that Sle_61320 might not be involved in antibiotic transport and hence should not be part of the chaxalactin biosynthesis gene cluster.

According to the NCBI CDD search (Marchler-Bauer et al., 2015), Sle_61350 and Sle_61360 are related to carbohydrate and amino acid transport and both are ABC transporter-domain containing proteins. The participation of these two enzymes and Sle_61320 in the transport of chaxalactins needs to be experimentally determined.

3.5.3 Metabolic engineering strategies for increased chaxalactin production by *S. leeuwenhoekii* C34

Chaxalactins A−C are new antibiotics produces by *S. leeuwenhoekii* C34 that displayed strong activity against Gram-positive bacteria, with MIC values of <1 µg/ml against *Staph. aureus* and 3-6 µg/ml against *L. monocytogenes* and *B. subtilis* (Rateb et al., 2011b). Other polyketides produced by *S. leeuwenhoekii* C34 are the chaxamycins. Chaxamycin A is the main fermentation product of *S. leeuwenhoekii* C34 in modified ISP2 liquid medium and in many other culture media (Rateb et al., 2011b). The polyketide backbone of chaxamycin A is synthesised from extensor units derived of malonyl-CoA and (2*S*)-methylmalonyl-CoA, which are the same as those used in chaxalactin biosynthesis. This implies that there is competition for precursor molecules that could be alleviated by deleting the 27 genes that are part of the chaxamycin biosynthesis gene cluster, located within nt position 1,210,347 to 1,290,550 in the genome of *S. leeuwenhoekii* C34.

Other biosynthetic routes like those for desferrioxamine E and hygromycin A, also compete for metabolites derived from primary metabolism. Therefore, the deletion of the locus of the desferrioxamine E biosynthesis gene cluster, which is between nt 5,237,176 to 5,244,356 (*sle_44550* to *sle_44600*) is proposed. In addition, deletion of the hygromycin A biosynthesis gene cluster that is located from nt 160,425 to 189,028 (*sle_01610* to *sle_01870*) is also proposed.

3.6 Conclusions

1. The domain architecture of subunits of the PKS: Sle_61410 (CxlA), Sle_61430 (CxlB), Sle_61440 (CxlC), Sle_61450 (CxlD) and Sle_61460 (CxlE), is consistent with the structure of chaxalactin A and, therefore, is related to chaxalactin biosynthesis.

2. A stretch of genomic DNA of 80.7 kb (nt 7,146,903 to 7,227,608) contains the majority of the genes to support chaxalactin A−C biosynthesis in *S. leeuwenhoekii* C34.

Chapter 4

Wrapping-up: general conclusions and perspectives of this thesis

In this thesis, three major achievements were accomplished: the development of genetic manipulation procedures of *S. leeuwenhoekii* C34, the identification of the chaxamycin biosynthesis gene cluster and the proposition of the putative locus of the genes related to chaxalactin biosynthesis. In addition, questions related to the biosynthesis of both chaxalactins and chaxamycins and other related to the genetics of *S. leeuwenhoekii* C34, remain to be answered in future work.

The first challenge encountered in this work was the assessment of the use of molecular tools for the genetic manipulation of *S. leeuwenhoekii* C34 for the identification of biosynthesis gene clusters and for future improvement of chaxamycin and chaxalactin production by metabolic engineering strategies, which require the deletion or over-expression of genes. Many plasmids are used for routine cloning in actinomycetes and, some of them use the integration systems from bacteriophages C31 and BT1. Bioinformatic analysis of the genome of *S. leeuwenhoekii* C34 revealed the presence of the canonical *attB* sites recognised by the integrases of bacteriophages C31 and BT1, being this the first evidence that the use of these site specific integration plasmids was feasible in *S. leeuwenhoekii* C34. Another enquiry was whether the methylation-deficient plasmid transfer system, *E. coli* ET12567/pUZ8002, would effectively mobilise plasmid DNA into *S. leeuwenhoekii* C34, and whether the antibiotics commonly used to select for genetically modified bacteria do effectively kill *S. leeuwenhoekii* C34 wild-type. So, firstly, the antimicrobial susceptibility of *S. leeuwenhoekii* C34 to thiostrepton, apramycin, hygromycin B and kanamycin was experimentally verified and its resistance to nalidixic acid, which is used to kill *E. coli* strains, was shown.

In this context, plasmid transfer from *E. coli* ET12567/pUZ8002 into *S. leeuwenhoekii* C34 was experimentally demonstrated using pSET152 that carries both an apramycin resistance gene and an integration system that recognises the ΦC31 *attB* site of *S. leeuwen-*

hoekii C34. Apramycin resistant exconjugants were observed and colony PCR confirmed that the integration of pSET152 in its genome was achieved. This result demonstrates that *S. leeuwenhoekii* C34 was able to accept foreign DNA from the *E. coli* methylation-deficient system. In addition, it remains to be elucidated whether *S. leeuwenhoekii* C34 has a methylation-restriction system as usually found in *Streptomyces* strains. Three putative methylation restriction systems were found in the genome of *S. leeuwenhoekii* C34: Sle_04470 that has an Mrr_cat domain, and Sle_34180 and Sle_45620 that contain a HNH_c domain. These protein sequences were homologous to those methylation-restriction systems studied in *S. coelicolor*, which recognise Dam- and Hsd- methylation patterns in foreign DNA: Sco4213, Sco4631 and Sco3262, respectively.

The effect of the presence of divalent chloride salts in the conjugation medium was another topic covered in this thesis. The individual effect of $MgCl_2$ and $CaCl_2$ on the conjugation frequency (number of exconjugants per number of donor cells) was assessed between *E. coli* ET12567/pUZ8002/pSET152 and *S. leeuwenhoekii* C34. Either, 120 mM $MgCl_2$ or 60 mM $CaCl_2$ significantly increased the conjugation frequency compared to no salt addition. Thus, it makes sense that conjugation performance could be further improved by studying the effect of the interaction of different salt concentrations in the conjugation medium. However, the underlying mechanism that explains this improvement of conjugation performance by the presence of divalent chloride salts is not clear.

De novo sequencing of the entire genome of *S. leeuwenhoekii* C34 was carried out, combining Illumina MiSeq and PacBio RS II SMRT sequencing technologies. Illumina MiSeq is a well-established technology that yields sequencing reads of several hundred bp, with very high accuracy, whereas PacBio is a relatively new sequencing platform that yields sequencing reads of several thousand bp, with a higher error rate than that obtained with Illumina. At the beginning of this work, a draft version of the genome of *S. leeuwenhoekii* C34 was kindly provided by Giles van Wezel and Geneviève Girard (University of Leiden, The Netherlands), with which the first genome mining analyses were carried out. They obtained the genome sequence assembled in 658 contigs, by using Illumina/-Solexa 100-nt paired-end-reads. One limitation of this version of the genome was the average contig size of the assembly, which was 12,546 bp, making difficult the identification of gene clusters that were larger than 80 kb, such as those of chaxamycins and chaxalactins.

The genomes of actinomycetes are difficult to assemble due to the presence of highly repetitive sequences, especially in those regions that encode for PKSs or NRPSs. Therefore, it makes sense that the combination of both sequencing technologies, Illumina MiSeq and PacBio, would take advantage of their intrinsic features to obtain a more reliable sequence of the genome of *S. leeuwenhoekii* C34 than that obtained if each of these sequencing technologies were used in a separate fashion. Long sequencing reads of PacBio were used to obtain longer contig length and then, the highly accurate short sequencing reads of Illumina MiSeq were used to correct the PacBio assembly. Thus, using this approach, the genome of *S. leeuwenhoekii* C34 was obtained as a single contig, with accuracy higher than 99% (0.03768% error). This type of procedure should be standard for sequencing projects of actinomycetes that aim to identify large biosynthesis gene clusters

with a high level of confidence.

The second challenge encountered in this work was the identification of the chaxamycin biosynthesis gene cluster, which was carried out using a genome mining approach that involved *in silico* identification of putative genes, heterologous expression and gene deletion. The chemical structures of chaxamycins A−D aided the identification of a stretch of genomic DNA of 27 genes that spanned 80.2 kb. This region contained genes devoted to the biosynthesis of the starter molecule of chaxamycins, AHBA, 5 PKS subunits and other genes encoding for enzymes related to tailoring reactions of the polyketide, consistent with those required to explain chaxamycin biosynthesis.

It is important to mention that, in addition to finding the genes related to the biosynthesis of chaxamycins, synteny analysis and local alignments were carried out to propose putative ends of the chaxamycin biosynthesis gene cluster. Synteny analysis is a widely used strategy that aims at the identification of highly conserved regions that flank a unique stretch of DNA, which usually indicates the presence of a gene cluster. Local alignments of a putative gene cluster are performed against other gene clusters that synthesise similar compounds, which in this case were the rifamycin, naphthomycin and saliniketal biosynthesis gene clusters. This approach looks for highly conserved DNA regions, complementing the synteny analysis towards the identification of the ends of the cluster. Thus, these strategies allowed the identification of a putative chaxamycin biosynthesis gene cluster of 80.2 kb, mentioned earlier.

For the heterologous expression experiment, a library of *S. leeuwenhoekii* C34 genomic DNA was constructed, which was later screened for the proposed ends of the chaxamycin biosynthesis gene cluster. This resulted in a vector, pIJ12853, that contained an insert of 145 kb of genomic DNA of *S. leeuwenhoekii* C34, which included the 80.2 kb region (the 27 genes). pIJ12853 was cloned into *S. coelicolor* M1152 (strain M1650), leading to the production of all chaxamycin species. This shows that the genes necessary for chaxamycin biosynthesis were present in the 145 kb DNA insert, and that *S. coelicolor* M1152 was able to produce chaxamycins.

It cannot be ruled out that genes of the heterologous host could be taking over biosynthesis reactions or transport steps in the chaxamycin biosynthesis and therefore, two important aspects of this result remain to be clarified in future work. First, the 145 kb-insert of pIJ12853 comprises 93 predicted coding sequences (PCS), of which 27 genes are proposed as the chaxamycin biosynthesis gene cluster. Analysis of synteny and local alignments strongly suggested that these 27 genes were related to chaxamycin biosynthesis. However, out of the remaining 66 PCSs, only 2 could be involved in AHBA synthesis: Sle_09700 and Sle_09710, which are outside of the 80.2 kb DNA region. These two proteins are predicted as units of a transketolase enzyme, which are homologous to those present in the rifamycin and saliniketal biosynthesis gene clusters, and at least in the saliniketal biosynthesis the deletion of those homologous led to the abolition of AHBA biosynthesis and saliniketals while this did not occur in the rifamycin biosynthesis. Thus, it needs to be elucidated whether these two *S. leeuwenhoekii* C34 proteins are related to

AHBA biosynthesis.

The second aspect that remains to be clarified in future work is the experimental verification of the ends of the chaxamycin biosynthesis gene cluster. This could be carried out using the pIJ12853, *S. coelicolor* M1152 and PCR-targeting. PCR-targeting consists of the replacement of a chromosomal sequence within a genomic insert in an artificial chromosome or a cosmid of an actinomycete by a selectable marker that has been generated by PCR using primers with 39 nt homology to the region to be replaced. The replacement of the gene of interest by the selectable marker is performed using the λ RED system of *E. coli*. Then the selectable marker could be excised from the vector by the FLP recombinase system, leaving a scar of 81 bp in the position where the replacement occurred. The regions that flank the 80.2 kb in pIJ12853 could be sequentially replaced by a selectable marker, then excised and finally transferred to *S. coelicolor* M1152 to assess chaxamycin production.

Another piece of important evidence that shows that the chaxamycin biosynthesis gene cluster was identified in the producer strain was the deletion of the AHBA synthase gene, *cxmK*, which led to the abolition of chaxamycin production by the strain M1653. This was carried out to demonstrate that *cxmK* gene is essential for AHBA and chaxamycin biosynthesis. Chemical complementation with commercial AHBA restored the production of all chaxamycin species in *S. leeuwenhoekii* M1653, which implies that the chaxamycin PKS was able to use exogenous sources of the precursor molecule to initiate the biosynthesis of chaxamycins. Thus, it makes sense, that production of unnatural chaxamycins is feasible with the strain M1653, by feeding it with analogues of AHBA that, for example, contain a chlorine or a fluorine atom in the C-4 position of AHBA. This position is important in ansamycin-type antibiotics, since it is usually modified by a methylation (*e. g.* chaxamycins, streptovaricins) or chlorination (*e. g.* naphthomycins) reaction, making this a possible target for modification. In addition, this kind of experiments must be correlated with bioassays, aiming to obtain improved biological activity of these unnatural chaxamycins compared to the ones produced by *S. leeuwenhoekii* C34 wild-type.

In addition, several targets for metabolic engineering have been proposed in this thesis and they remain to be experimentally demonstrated that they actually increase chaxamycin production in *S. leeuwenhoekii* C34. Thus, the chaxamycin biosynthesis gene cluster leaves many open questions that require an intense workload to be answered. Some of the open questions are: what is the role of regulatory genes *cxmY* and *cxmZ* in the biosynthesis of chaxamycins? Which step of the chaxamycin biosynthesis is catalysed by the cytochromes P450 encoded in the cluster? What is the locus of a chaxamycin transport system in the genome of *S. leeuwenhoekii* C34? Among others.

The third challenge encountered in this thesis was the bioinformatic identification of the putative biosynthetic genes that form the chaxalactin biosynthesis gene cluster of *S. leeuwenhoekii* C34. This bioinformatic search was aided by the chemical structure of chaxalactins A−C, which allowed formulation of predictions such as the number of modules and the minimum number of domains that the PKS should need and the possible tailoring

reactions involved in the production of chaxalactin B and chaxalactin C. From the results, it was possible to conclude that five genes that encoded for PKS subunits (*sle_61410*, *sle_61430*, *sle_61440*, *sle_61450* and *sle_61460*), had a protein domain architecture consistent with that predicted for the biosynthesis of chaxalactins. In addition, chaxalactin A was proposed as the unique product of the synthesis of the PKS, since its whole structure, contrary to chaxalactin B and chaxalactin C, obeys the logic of biosynthesis carried out by a type I PKS. The structure of chaxalactin B and chaxalactin C are likely to be products of post-synthetic modifications.

Several genes that encoded for cytochrome P450, Rieske domain-containing proteins and methyltransferases were found in proximity to the PKS genes. In particular, one gene, *sle_61400* (*cxlF*), encoded for a cytochrome P450 which seems to be translationally coupled to the first PKS subunit, *sle_61410* (*cxlA*). This cytochrome P450 possesses almost all conserved motifs found in a typical cytochrome P450, but it seems to lack the absolutely conserved cysteine residue that coordinates the haem group. At first glance, *cxlF* would encode for an inactive cytochrome P450, but protein alignment with other cytochromes P450, from other organisms that also lack the conserved cysteine, suggests that there might be an exception to the rule. Downstream to *sle_61460* (*cxlE*) there are two genes that encode for Rieske domain-containing proteins, *sle_61490* and *sle_61500*. Rieske proteins are known to introduce di-hydroxylation to the substrate where they act. Therefore, the participation of *cxlF*, *sle_61490* and *sle_61500* in chaxalactin B biosynthesis remain to be experimentally determined. Two genes, *sle_61390* and *sle_61510*, encoded for active *O*-methyltransferases, since they possess all the conserved motifs found in active methyltransferases. These are likely to participate in the conversion of chaxalactin B into chaxalactin C but this hypothesis needs to be verified experimentally.

The results presented in this thesis described the process aimed at the identification of the biosynthetic genes involved in the production of chaxamycins and chaxalactins by the novel species, *S. leeuwenhoekii* C34. Bioinformatic-based methodologies, combined with wet-lab verifications were carried out to identify the chaxamycin biosynthesis gene cluster. The establishment of genetic modification procedures of *S. leeuwenhoekii* C34 has been reported and largely discussed in this thesis. The proposition of the locus of the putative chaxalactin biosynthesis gene cluster was also presented and supported with recent knowledge in the area. Thus, the significance of this thesis is the use of different methodologies to exploit the potential of a microorganism producer of new antibiotics, which are urgently needed because of the recurrent emergence of antibiotic-resistant bacteria with concomitant outbreaks. Chile has a variety of environments and climates that harbour an immense microbial diversity that are the potential producers of novel antibiotics, antifungal or anticancer drugs as well as novel metabolites with other useful properties. For this reason, more efforts in preserving ecosystems and bioprospecting initiatives should be made to protect and study our microbial resources in a sustainable way.

Bibliography

Absalón, A. E., Fernández, F. J., Olivares, P. X., Barrios-González, J., Campos, C., and Mejía, A. (2007). RifP; a membrane protein involved in rifamycin export in *Amycolatopsis mediterranei*. *Biotechnology Letters*, 29(6):951–958.

Admiraal, S. J., Khosla, C., and Walsh, C. T. (2002). The loading and initial elongation modules of rifamycin synthetase collaborate to produce mixed aryl ketide products. *Biochemistry*, 41(16):5313–5324.

Admiraal, S. J., Walsh, C. T., and Khosla, C. (2001). The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. *Biochemistry*, 40(20):6116–6123.

Allenby, N. E. E., Laing, E., Bucca, G., Kierzek, A. M., and Smith, C. P. (2012). Diverse control of metabolism and other cellular processes in *Streptomyces coelicolor* by the PhoP transcription factor: genome-wide identification of *in vivo* targets. *Nucleic Acids Research*, 40(19):9543–9556.

Alting-Mees, M. A. and Short, J. M. (1989). pBluescript II: gene mapping vectors. *Nucleic Acids Research*, 17(22):9494.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402.

Ansari, M., Sharma, J., Gokhale, R. S., and Mohanty, D. (2008). *In silico* analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinformatics*, 9(1):454.

Anton, B. P. and Raleigh, E. A. (2004). Transposon-mediated linker insertion scanning mutagenesis of the *Escherichia coli* McrA endonuclease. *Journal of Bacteriology*, 186(17):5699–5707.

August, P. R., Tang, L., Yoon, Y. J., Ning, S., Müller, R., Yu, T. W., Taylor, M., Hoffmann, 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. *Chemistry & Biology*, 5(2):69–79.

Austin, M. B. and Noel, J. P. (2003). The chalcone synthase superfamily of type iii polyketide synthases. *Natural Product Reports*, 20(1):79–110.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9(1):75.

Baltz, R. H. (2008). Renaissance in antibacterial discovery from actinomycetes. *Current opinion in pharmacology*, 8(5):557–563.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885):141–147.

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of antibiotics*, 58(1):1.

Bevitt, D. J., Cortes, J., Haydock, S. F., and Leadlay, 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. *European Journal of Biochemistry*, 204(1):39–49.

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current opinion in microbiology*, 8(2):208–215.

Bibb, M. J., Findlay, P. R., and Johnson, M. W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene*, 30(1-3):157–166.

Bibb, M. J., Janssen, G. R., and Ward, J. M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene*, 38(1-3):215–226.

Bierman, M., Logan, R., O'brien, K., Seno, E. T., Rao, R. N., and Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*, 116:43–49.

Bisang, C., Long, P. F., Cortés, J., Westcott, J., Crosby, J., Matharu, A. L., Cox, R. J., Simpson, T. J., Staunton, J., and Leadlay, P. F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, 401(6752):502–505.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, 41(Web Server issue):W204–W212.

Bologa, C. G., Ursu, O., Oprea, T. I., Melançon III, C. E., and Tegos, G. P. (2013). Emerging trends in the discovery of natural product antibacterials. *Current Opinion in Pharmacology*, 13(5):678–687.

Bujnicki, J. M. and Rychlewski, L. (2001). Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene*, 267(2):183–191.

Bull, A. T. and Asenjo, J. A. (2013). Microbiology of hyper-arid environments: recent insights from the Atacama Desert, Chile. *Antonie van Leeuwenhoek*, 103(6):1173–1179.

Busarakam, K., Bull, A. T., Girard, G., Labeda, D. P., van Wezel, G. P., and Goodfellow, M. (2014). *Streptomyces leeuwenhoekii* sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in *Streptomyces* gene trees. *Antonie van Leeuwenhoek*, 105(5):849–861.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, 104(6):901–912.

Capyk, J. K., D'Angelo, I., Strynadka, N. C., and Eltis, L. D. (2009). Characterization of 3-ketosteroid 9α-hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 284(15):9937–9946.

Centers for Disease Control and Prevention (2013). Antimicrobial resistance threats in the United States, 2013.

Chater, K. F. (1974). Rifampicin-resistant mutants of *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, 80(1):277–290.

Chen, W., Lee, M.-K., Jefcoate, C., Kim, S.-C., Chen, F., and Yu, J.-H. (2014). Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution*, 6(7):1620–1634.

Choi, S.-U., Lee, C.-K., Hwang, Y.-I., Kinoshita, H., and Nihira, T. (2004). Intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to *Kitasatospora setae*, a bafilomycin B1 producer. *Archives of Microbiology*, 181(4):294–298.

Claxton, H. B., Akey, D. L., Silver, M. K., Admiraal, S. J., and Smith, J. L. (2009). Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *Journal of Biological Chemistry*, 284(8):5021–5029.

Davies, J. and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 74(3):417–433.

Davies, J. and O'Connor, S. (1978). Enzymatic modification of aminoglycoside antibiotics: 3-*N*-acetyltransferase with broad specificity that determines resistance to the novel aminoglycoside apramycin. *Antimicrobial Agents and Chemotherapy*, 14(1):69–72.

Demain, A. L. and Fang, A. (2000). The natural functions of secondary metabolites. *Advances in biochemical engineering/biotechnology*, 69:1–39.

Demydchuk, J., Oliynyk, Z., and Fedorenko, V. (1998). Analysis of a kanamycin resistance gene (*kmr*) from *Streptomyces kanamyceticus* and a mutant with increased aminoglycoside resistance. *Journal of basic microbiology*, 38(4):231–239.

Dewick, P. M. (2009). *Medicinal Natural Products A Biosynthetic Approach 3rd Edition*. John Wiley & Sons Ltd.

Dhote, V., Gupta, S., and Reynolds, K. A. (2008). An *O*-phosphotransferase catalyzes phosphorylation of hygromycin A in the antibiotic-producing organism *Streptomyces hygroscopicus*. *Antimicrobial Agents and Chemotherapy*, 52(10):3580–3588.

Drlica, K. and Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiology and Molecular Biology Reviews*, 61(3):377–392.

Du, L., Liu, R.-H., Ying, L., and Zhao, G.-R. (2012). An efficient intergeneric conjugation of DNA from *Escherichia coli* to mycelia of the lincomycin-producer *Streptomyces lincolnensis*. *International journal of molecular sciences*, 13(4):4797–4806.

Dunn, B. J., Cane, D. E., and Khosla, C. (2013). Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry*, 52(11):1839–1841.

Elliot, M. A. and Flärdh, K. (2001). *Streptomycete spores*. eLS. John Wiley & Sons, Ltd, Chichester, UK.

Elsayed, S. S., Trusch, F., Deng, H., Raab, A., Prokes, I., Busarakam, K., Asenjo, J. A., Andrews, B. A., van West, P., Bull, A. T., Goodfellow, M., Yi, Y., Ebel, R., Jaspars, M., and Rateb, M. E. (2015). Chaxapeptin, a lasso peptide from extremotolerant *Streptomyces leeuwenhoekii* strain C58 from the hyperarid Atacama Desert. *The Journal of organic chemistry*.

Energy and Commerce Committee, Health Subcommittee United States House of Representatives (2014). Written statement of Barbara E. Murray, MD, FIDSA, president infectious diseases society of America; 21st century cures: examining ways to combat antibiotic resistance and foster new drug development.

Fernández-Moreno, M. A., Martínez, E., Boto, L., Hopwood, D. A., and Malpartida, F. (1992). Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *Journal of Biological Chemistry*, 267(27):19278–19290.

Ferraro, D. J., Gakhar, L., and Ramaswamy, S. (2005). Rieske business: structure-function of Rieske non-heme oxygenases. *Biochemical and biophysical research communications*, 338(1):175–190.

Flärdh, K. and Buttner, M. J. (2009). *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*, 7(1):36–49.

Flatt, P. M. and Mahmud, T. (2007). Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Natural Product Reports*, 24(2):358–392.

Floss, H. G. and Yu, T.-W. (1999). Lessons from the rifamycin biosynthetic gene cluster. *Current Opinion in Chemical Biology*, 3(5):1–6.

Floss, H. G. and Yu, T.-W. (2005). Rifamycin-mode of action, resistance, and biosynthesis. *Chemical Reviews*, 105(2):621–632.

Floss, H. G., Yu, T.-W., and Arakawa, K. (2011). The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC_7N units in ansamycin and mitomycin antibiotics: a review. *The Journal of Antibiotics*, 64(1):35–44.

Foulston, L. C. and Bibb, M. J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13461–13466.

Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology*, 13(2):150–159.

García Angulo, V. A., Bonomi, H. R., Posadas, D. M., Serer, M. I., Torres, A. G., Zorreguieta, Á., and Goldbaum, F. A. (2013). Identification and characterization of RibN, a novel family of riboflavin transporters from *Rhizobium leguminosarum* and other proteobacteria. *Journal of Bacteriology*, 195(20):4611–4619.

Garg, A., Xie, X., Keatinge-Clay, A., Khosla, C., and Cane, D. E. (2014). Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange. *Journal of the American Chemical Society*, 136(29):10190–10193.

Gomez-Escribano, J. P. and Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology*, 4(2):207–215.

Gomez-Escribano, J. P. and Bibb, M. J. (2014). Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *Journal of Industrial Microbiology & Biotechnology*, 41(2):425–431.

Gomez-Escribano, J. P., Castro, J. F., Razmilic, V., Chandra, G., Andrews, B., Asenjo, J. A., and Bibb, M. J. (2015). The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics*, 16(1):485.

Gomez-Escribano, J. P., Song, L., Fox, D. J., Yeo, V., Bibb, M. J., and Challis, G. L. (2012). Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the *cpk* gene cluster of *Streptomyces coelicolor* M145. *Chemical Science*, 3(9):2716–2720.

González-Cerón, G., Miranda-Olivares, O. J., and Servín-González, L. (2009). Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiology Letters*, 301(1):35–43.

Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences*, 87(12):4645–4649.

Gregory, M. A., Till, R., and Smith, M. C. M. (2003). Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *Journal of Bacteriology*, 185(17):5320–5323.

Guo, J. and Frost, J. W. (2002). Kanosamine biosynthesis: a likely source of the aminoshikimate pathway's nitrogen atom. *Journal of the American Chemical Society*.

Ha, J. Y., Min, J. Y., Lee, S. K., Kim, H. S., Kim, D. J., Kim, K. H., Lee, H. H., Kim, H. K., Yoon, H.-J., and Suh, S. W. (2006). Crystal structure of 2-nitropropane dioxygenase complexed with FMN and substrate. Identification of the catalytic base. *Journal of Biological Chemistry*, 281(27):18660–18667.

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular biology and evolution*, 30(5):1229–1235.

Hamano, Y., Hoshino, Y., Nakamori, S., and Takagi, H. (2004). Overexpression and characterization of an aminoglycoside 6´-*N*-acetyltransferase with broad specificity from an ε-poly-L-lysine producer, *Streptomyces albulus* IFO14147. *Journal of Biochemistry*, 136(4):517–524.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, 166(4):557–580.

Hara, O. and Hutchinson, C. R. (1992). A macrolide 3-*O*-acyltransferase gene from the midecamycin-producing species *Streptomyces mycarofaciens*. *Journal of Bacteriology*, 174(15):5141–5144.

Hayashi, K., Ogino, K., Oono, Y., Uchimiya, H., and Nozaki, H. (2001). Yokonolide A, a new inhibitor of auxin signal transduction, from *Streptomyces diastatochromogenes* B59. *The Journal of Antibiotics*, 54(7):573–581.

Herbert, R. B. (1989). *The biosynthesis of secondary metabolites*. Springer.

Hobbs, G., Frazer, C. M., Gardner, D. C. J., Cullum, J. A., and Oliver, S. G. (1989). Dispersed growth of *streptomyces* in liquid culture. *Applied Microbiology and Biotechnology*, 31:272–277.

Hong, H.-J., Hutchings, M. I., Hill, L. M., and Buttner, M. J. (2005). The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *Journal of Biological Chemistry*, 280(13):13055–13061.

Hopwood, D. A. and Wright, H. M. (1978). Bacterial protoplast fusion: recombination in fused protoplasts of *Streptomyces coelicolor*. *Molecular & General Genetics MGG*, 162(3):307–317.

Huang, G., Zhang, L., and Birch, R. G. (2001). A multifunctional polyketide–peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiology*, 147(3):631–642.

Huang, T.-W., Hsu, C.-C., Yang, H.-Y., and Chen, C. W. (2013). Topoisomerase IV is required for partitioning of circular chromosomes but not linear chromosomes in *Streptomyces*. *Nucleic Acids Research*, 41(22):10403–10413.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1):119.

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S., and Fujita, N. (2013). DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, 41(Database issue):D408–D414.

Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., and Ōmura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proceedings of the National Academy of Sciences*, 96(17):9509–9514.

Illumina Inc. (2010). Illumina sequencing technology. Technical report.

Instituto de Salud Pública de Chile (2013a). Vigilancia de *Enterococcus* spp. resistente a vancomicina chile, 2010–2012. *Boletín Instituto de Salud Pública de Chile*, 3(10):1–21.

Instituto de Salud Pública de Chile (2013b). Vigilancia de *Staphylococcus aureus* meticilina resistente adquirido en la comunidad. Chile, 2007, 2012. *Boletín Instituto de Salud Pública de Chile*, 3(7):1–21.

Jack, D. L., Yang, N. M., and Saier, M. H. (2001). The drug/metabolite transporter superfamily. *European Journal of Biochemistry*, 268(13):3620–3639.

Jacoby, G. A. (2005). Mechanisms of resistance to quinolones. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 41 Suppl 2(Supplement 2):S120–6.

Jones, A. C., Gust, B., Kulik, A., Heide, L., Buttner, M. J., and Bibb, M. J. (2013). Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PLoS ONE*, 8(7):e69319.

Kagan, R. M. and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Archives of biochemistry and biophysics*, 310(2):417–427.

Kang, Q., Shen, Y., and Bai, L. (2012). Biosynthesis of 3,5-AHBA-derived natural products. *Natural Product Reports*, 29(2):243–263.

Kapur, S., Lowry, B., Yuzawa, S., Kenthirapalan, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2012). Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(11):4110–4115.

Keatinge-Clay, A. T. (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chemistry & Biology*, 14(8):898–908.

Keatinge-Clay, A. T. (2012). The structures of type I polyketide synthases. *Natural Product Reports*, 29(10):1050–1073.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical Streptomyces manual*. John Innes Foundation, Norwich, UK, second edition.

Kihara, T., Kusakabe, H., Nakamura, G., Sakurai, T., and Isono, K. (1981). Cytovaricin, a novel antibiotic. *The Journal of Antibiotics*, 34(8):1073–1074.

Kim, B. S., Sherman, D. H., and Reynolds, K. A. (2004). An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. *Protein engineering, design & selection : PEDS*, 17(3):277–284.

Kim, H.-J., Karki, S., Kwon, S.-Y., Park, S.-H., Nahm, B.-H., Kim, Y.-K., and Kwon, H.-J. (2014). A single module type I polyketide synthase directs *de novo* macrolactone biogenesis during galbonolide biosynthesis in *Streptomyces galbus*. *Journal of Biological Chemistry*, 289(50):34557–34568.

Kliebenstein, D. J. and Osbourn, A. (2012). Making new molecules – evolution of pathways for novel metabolites in plants. *Current opinion in plant biology*, 15(4):415–423.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a Hidden Markov Model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580.

Kumar, S., Mukherjee, M. M., and Varela, M. F. (2013). Modulation of bacterial multidrug resistance efflux pumps of the major facilitator superfamily. *International journal of bacteriology*, 2013.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M., and Hertweck, C. (2010). Functionally distinct modules operate two consecutive $\alpha,\beta\rightarrow\beta,\gamma$ double-bond shifts in the rhizoxin polyketide assembly line. *Angewandte Chemie (International ed. in English)*, 49(8):1460–1464.

Kwan, D. H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J. C. C., Haydock, S. F., and Leadlay, P. F. (2008). Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chemistry & Biology*, 15(11):1231–1240.

Lamb, D. C. and Waterman, M. R. (2013). Unusual properties of the cytochrome P450 superfamily. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1612):20120434.

Letunic, I. and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39(Web Server issue):W475–8.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research*, 43(Database issue):D257–60.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5):589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, S., Li, Y., Lu, C., Zhang, J., Zhu, J., Wang, H., and Shen, Y. (2015). Activating a cryptic ansamycin biosynthetic gene cluster to produce three new naphthalenic octaketide ansamycins with *n*-pentyl and *n*-butyl side chains. *Organic Letters*, 17(15):3706–3709.

Li, S., Wang, H., Li, Y., Deng, J., Lu, C., Shen, Y., and Shen, Y. (2014). Biosynthesis of hygrocins, antitumor naphthoquinone ansamycins produced by *Streptomyces* sp. LZ35. *Chembiochem*, 15(1):94–102.

Liu, C., Zhang, J., Lu, C., and Shen, Y. (2015). Heterologous expression of galbonolide biosynthetic genes in *Streptomyces coelicolor*. *Antonie van Leeuwenhoek*, 107(5):1359–1366.

Liu, G., Chater, K. F., Chandra, G., Niu, G., and Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiology and molecular biology reviews : MMBR*, 77(1):112–143.

Long, P. F., Wilkinson, C. J., Bisang, C. P., Cortes, J., Dunster, N., Oliynyk, M., McCormick, E., McArthur, H., Méndez, C., Salas, J. A., Staunton, J., and Leadlay, P. F. (2002). Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Molecular Microbiology*, 43(5):1215–1225.

Lovering, A. M., White, L. O., and Reeves, D. S. (1987). AAC(1): a new aminoglycoside-acetylating enzyme modifying the C1 aminogroup of apramycin. *Journal of Antimicrobial Chemotherapy*, 20(6):803–813.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H., and MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, 111(1):61–68.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43(Database issue):D222–D226.

Martin, J. F. and Liras, P. (1989). Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annual Reviews in Microbiology*, 43(1):173–206.

McCoy, J. G., Johnson, H. D., Singh, S., Bingman, C. A., Lei, I.-K., Thorson, J. S., and Phillips, G. N. (2009). Structural characterization of CalO2: a putative orsellinic acid P450 oxidase in the calicheamicin biosynthetic pathway. *Proteins*, 74(1):50–60.

McMurry, J. (2011). *Organic Chemistry with Biological Applications 2e*. Mary Finch.

Medema, M. H., Breitling, R., and Takano, E. (2011). Synthetic biology in *Streptomyces* bacteria. *Methods in enzymology*, 497:485–502.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J. A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., and Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, 43(Database issue):D213–21.

Murakami, T., Holt, T. G., and Thompson, C. J. (1989). Thiostrepton-induced gene expression in *Streptomyces lividans*. *Journal of Bacteriology*, 171(3):1459–1466.

Murakami, T., Nojiri, C., Toyama, H., Hayashi, E., Katumata, K., Anzai, H., Matsuhashi, Y., Yamada, Y., and Nagaoka, K. (1983). Cloning of antibiotic-resistance genes in *Streptomyces*. *The Journal of Antibiotics*, 36(10):1305–1311.

Muth, G., Nußbaumer, B., Wohlleben, W., and Pühler, A. (1989). A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. *Molecular & General Genetics MGG*, 219(3):341–348.

Myronovskyi, M., Welle, E., Fedorenko, V., and Luzhetskyy, A. (2011). Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Applied and Environmental Microbiology*, 77(15):5370–5383.

Nachtigall, J., Kulik, A., Helaly, S., Bull, A. T., Goodfellow, M., Asenjo, J. A., Maier, A., Wiese, J., Imhoff, J. F., Süssmuth, R. D., and Fiedler, H.-P. (2011). Atacamycins A–C, 22-membered antitumor macrolactones produced by *Streptomyces* sp. C38. *The Journal of Antibiotics*, 64(12):775–780.

Nett, M. (2014). Genome mining: concept and strategies for natural product discovery. *Progress in the chemistry of organic natural products*, 99:199–245.

Nett, M. and König, G. M. (2007). The chemistry of gliding bacteria. *Natural Product Reports*, 24(6):1245–1261.

Novakova, R., Homerova, D., Feckova, L., and Kormanec, J. (2005). Characterization of a regulatory gene essential for the production of the angucycline-like polyketide antibiotic auricin in *Streptomyces aureofaciens* CCM 3239. *Microbiology*, 151(Pt 8):2693–2706.

Okanishi, M., Suzuki, K., and Umezawa, H. (1974). Formation and reversion of streptomycete protoplasts: cultural condition and morphological study. *Journal of General Microbiology*, 80(2):389–400.

Okoro, C. K., Brown, R., Jones, A. L., Andrews, B. A., Asenjo, J. A., Goodfellow, M., and Bull, A. T. (2009). Diversity of culturable actinomycetes in hyper-arid soils of the atacama desert, chile. *Antonie van Leeuwenhoek*, 95(2):121–133.

Olano, C., Lombó, F., Méndez, C., and Salas, J. A. (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metabolic Engineering*, 10(5):281–292.

Ōmura, S., Tanaka, Y., Hisatome, K., Miura, S., Takahashi, Y., Nakagawa, A., Imai, H., and Woodruff, H. B. (1988). Phthoramycin, a new antibiotic active against a plant pathogen, *Phytophthora* sp. *The Journal of Antibiotics*, 41(12):1910–1912.

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in genetics : TIG*, 26(10):449–457.

Outterson, K. (2014). High-quality antibiotics. Technical report, Center for Disease Dynamics, Economics & Policy.

Outterson, K., Powers, J. H., Seoane-Vazquez, E., Rodriguez-Monguio, R., and Kesselheim, A. S. (2013). Approval and withdrawal of new antibiotics and other antiinfectives in the U.S., 1980-2009. 41(3):688–696.

Pacific Biosciences (2015). Revolutionize genomics with SMRT® sequencing single molecule, real-time technology.

Pardo, J. M., Malpartida, F., Rico, M., and Jiménez, A. (1985). Biochemical basis of resistance to hygromycin B in *Streptomyces hygroscopicus*–the producing organism. *Journal of General Microbiology*, 131(6):1289–1298.

Park, S. R., Yoo, Y. J., Ban, Y. H., and Yoon, Y. J. (2010). Biosynthesis of rapamycin and its regulation: past achievements and recent progress. *The Journal of Antibiotics*, 63(8):434–441.

Pawlik, K., Kotowska, M., Chater, K. F., Kuczek, K., and Takano, E. (2007). A cryptic type I polyketide synthase (*cpk*) gene cluster in *Streptomyces coelicolor* A3(2). *Archives of Microbiology*, 187(2):87–99.

Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery*, 6(1):29–40.

Peano, C., Damiano, F., Forcato, M., Pietrelli, A., Palumbo, C., Corti, G., Siculella, L., Fuligni, F., Tagliazucchi, G. M., De Benedetto, G. E., Bicciato, S., De Bellis, G., and Alifano, P. (2014). Comparative genomics revealed key molecular targets to rapidly convert a reference rifamycin-producing bacterial strain into an overproducer by genetic engineering. *Metabolic Engineering*, 26C:1–16.

Pieper, R., Luo, G., Cane, D. E., and Khosla, C. (1995). Remarkably broad substrate specificity of a modular polyketide synthase in a cell-free system. *Journal of the American Chemical Society*, 117(45):11373–11374.

Podust, L. M. and Sherman, D. H. (2012). Diversity of P450 enzymes in the biosynthesis of natural products. *Natural Product Reports*, 29(10):1251–1266.

Poehlsgaard, J. and Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. *Nature Reviews Microbiology*, 3(11):870–881.

Poralla, K., Muth, G., and Hartner, T. (2000). Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters*, 189(1):93–95.

Priyam, A., Woodcroft, B. J., Rai, V., and Wurm, Y. SequenceServer: BLAST search made easy http://sequenceserver.com. *in preparation*.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1):341.

Que, L. and Tolman, W. B. (2008). Biologically inspired oxidation catalysis. *Nature*, 455(7211):333–340.

Ramirez, M. S. and Tolmasky, M. E. (2010). Aminoglycoside modifying enzymes. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy*, 13(6):151–171.

Rateb, M. E., Houssen, W. E., Arnold, M., Abdelrahman, M. H., Deng, H., Harrison, W. T. A., Okoro, C. K., Asenjo, J. A., Andrews, B. A., Ferguson, G., Bull, A. T., Goodfellow, M., Ebel, R., and Jaspars, M. (2011a). Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. *Journal of Natural Products*, 74(6):1491–1499.

Rateb, M. E., Houssen, W. E., Harrison, W. T. A., Deng, H., Okoro, C. K., Asenjo, J. A., Andrews, B. A., Bull, A. T., Goodfellow, M., Ebel, R., and Jaspars, M. (2011b). Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment. *Journal of Natural Products*, 74(9):1965–1971.

Rausch, H. and Lehmann, M. (1991). Structural analysis of the actinophage φC31 attachment site. *Nucleic Acids Research*, 19(19):5187–5189.

Rebets, Y., Dutko, L., Ostash, B., Luzhetskyy, A., Kulachkovskyy, O., Yamaguchi, T., Nakamura, T., Bechthold, A., and Fedorenko, V. (2008). Function of lanI in regulation of landomycin A biosynthesis in *Streptomyces cyanogenus* S136 and cross-complementation studies with *Streptomyces* antibiotic regulatory proteins encoding genes. *Archives of Microbiology*, 189(2):111–120.

Reeves, A. R., Brikun, I. A., Cernota, W. H., Leach, B. I., Gonzalez, M. C., and Weber, J. M. (2007). Engineering of the methylmalonyl-CoA metabolite node of *Saccharopolyspora erythraea* for increased erythromycin production. *Metabolic Engineering*, 9(3):293–303.

Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D. V., Hutchinson, C. R., and McDaniel, R. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry*, 42(1):72–79.

Rodriguez-Garcia, A., Santamarta, I., Perez-Redondo, R., Martin, J. F., and Liras, P. (2006). Characterization of a two-gene operon *epeRA* involved in multidrug resistance in *Streptomyces clavuligerus*. *Research in microbiology*, 157(6):559–568.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945.

Ryu, Y.-G., Butler, M. J., Chater, K. F., and Lee, K. J. (2006). Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Applied and Environmental Microbiology*, 72(11):7132–7139.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

Schmitt-John, T. and Engels, J. W. (1992). Promoter constructions for efficient secretion expression in *Streptomyces lividans*. *Applied Microbiology and Biotechnology*, 36(4):493–498.

Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. *Trends in biochemical sciences*, 28(6):329–335.

Seipke, R. F. and Loria, R. (2009). Hopanoids are not essential for growth of *Streptomyces scabies* 87-22. *Journal of Bacteriology*, 191(16):5216–5223.

Sezutsu, H., Le Goff, G., and Feyereisen, R. (2013). Origins of P450 diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1612):20120428.

Shao, Z. H., Ren, S. X., Liu, X. Q., Xu, J., Yan, H., Zhao, G.-P., and Wang, J. (2015). A preliminary study of the mechanism of nitrate-stimulated remarkable increase of rifamycin production in *Amycolatopsis mediterranei* U32 by RNA-seq. *Microbial Cell Factories*, 14(1):75.

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7(2):285–295.

Sherwood, E. J. and Bibb, M. J. (2013). The antibiotic planosporicin coordinates its own production in the actinomycete *Planomonospora alba*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27):E2500–2509.

Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S., and Ochi, K. (1996). Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 178(24):7276–7284.

Shirling, E. B. and Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *International Journal of Systematic Bacteriology*, 16(3):313–340.

Shoichet, B. K. (2013). Drug discovery: Nature's pieces. *Nature Chemistry*, 5(1):9–10.

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2012). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1):D344–D347.

Sim, E., Sandy, J., Evangelopoulos, D., Fullam, E., Bhakta, S., Westwood, I., Krylova, A., Lack, N., and Noble, M. (2008). Arylamine N-acetyltransferases in mycobacteria. *Current drug metabolism*, 9(6):510–519.

Skeggs, P. A., Holmes, D. J., and Cundliffe, E. (1987). Cloning of aminoglycoside-resistance determinants from *Streptomyces tenebrarius* and comparison with related genes from other actinomycetes. *Journal of General Microbiology*, 133(4):915–923.

Smith, S. and Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports*, 24(5):1041–1072.

Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A. M., and Donadio, S. (2000). Artificial chromosomes for antibiotic-producing actinomycetes. *Nature Biotechnology*, 18(3):343–345.

Staden, R., Beal, K. F., and Bonfield, J. K. (1999). *The Staden Package, 1998*. Humana Press.

Staley, A. L. and Rinehart, K. L. (1991). Biosynthesis of the streptovaricins: 3-amino-5-hydroxybenzoic acid as a precursor to the *meta*-C_7N unit. *The Journal of Antibiotics*, 44(2):218–224.

Staunton, J. and Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. *Natural Product Reports*, 18(4):380–416.

Sun, F.-H., Luo, D., Shu, D., Zhong, J., and Tan, H. (2014). Development of an intergeneric conjugal transfer system for xinaomycins-producing *Streptomyces noursei* Xinao-4. *International journal of molecular sciences*, 15(7):12217–12230.

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10):1282–1288.

Sydor, P. K., Barry, S. M., Odulate, O. M., Barona-Gómez, F., Haynes, S. W., Corre, C., Song, L., and Challis, G. L. (2011). Regio- and stereodivergent antibiotic oxidative carbocyclizations catalysed by Rieske oxygenase-like enzymes. *Nature Chemistry*, 3(5):388–392.

Taft, F., Brünjes, M., Knobloch, T., Floss, H. G., and Kirschning, A. (2009). Timing of the $\triangle$10,12-$\triangle$11,13 double bond migration during ansamitocin biosynthesis in *Actinosynnema pretiosum*. *Journal of the American Chemical Society*, 131(11):3812–3813.

Takahashi, K., Yoshihara, T., and Kurosawa, K. (2005). Ushikulides A and B, immunosuppressants produced by a strain of *Streptomyces* sp. *The Journal of Antibiotics*, 58(6):420–424.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30(12):2725–2729.

Thompson, J., Schmidt, F., and Cundliffe, E. (1982). Site of action of a ribosomal RNA methylase conferring resistance to thiostrepton. *Journal of Biological Chemistry*, 257(14):7915–7917.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25(24):4876–4882.

Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C., and Stroud, R. M. (2001). Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proceedings of the National Academy of Sciences*, 98(26):14808–14813.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, 35(Web Server issue):W71–W74.

Vaishnav, P. and Demain, A. L. (2011). Unexpected applications of secondary metabolites. *Biotechnology advances*, 29(2):223–229.

Vakulenko, S. B. and Mobashery, S. (2003). Versatility of aminoglycosides and prospects for their future. *Clinical microbiology reviews*, 16(3):430–450.

Valenzano, C. R., You, Y.-O., Garg, A., Keatinge-Clay, A., Khosla, C., and Cane, D. E. (2010). Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *Journal of the American Chemical Society*, 132(42):14697–14699.

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., and Wishart, D. S. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research*, 33(Web Server issue):W455–W459.

Västermark, Å., Almén, M. S., Simmen, M. W., Fredriksson, R., and Schiöth, H. B. (2011). Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of *Viridiplantae*. *BMC evolutionary biology*, 11(1):123.

Walton, J. R. (1978). Apramycin, a new aminocyclitol antibiotic. I. *In vitro* microbiological studies. *Journal of Antimicrobial Chemotherapy*, 4(4):309–313.

Wang, H. X., Chen, Y. Y., Ge, L., Fang, T. T., Meng, J., Liu, Z., Fang, X. Y., Ni, S., Lin, C., Wu, Y. Y., Wang, M. L., Shi, N. N., He, H. G., Hong, K., and Shen, Y. M. (2013). PCR screening reveals considerable unexploited biosynthetic potential of ansamycins and a mysterious family of AHBA-containing natural products in actinomycetes. *Journal of Applied Microbiology*, 115(1):77–85.

Wang, L., Tian, X., Wang, J., Yang, H., Fan, K., Xu, G., Yang, K., and Tan, H. (2009). Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21):8617–8622.

Wang, X.-K. and Jin, J.-L. (2014). Crucial factor for increasing the conjugation frequency in *Streptomyces netropsis* SD-07 and other strains. *FEMS Microbiology Letters*, 357(1):99–103.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015). antiSMASH 3.0–a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(W1):W237–43.

Weissman, K. J. and Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*, 3(12):925–936.

Wiesmann, K. E., Cortés, J., Brown, M. J., Cutter, A. L., Staunton, J., and Leadlay, P. F. (1995). Polyketide synthesis *in vitro* on a modular polyketide synthase. *Chemistry & Biology*, 2(9):583–589.

Wilson, M. C., Gulder, T. A. M., Mahmud, T., and Moore, B. S. (2010). Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the *sare1259*-encoded cytochrome P450. *Journal of the American Chemical Society*, 132(36):12757–12765.

Witkowski, A., Joshi, A. K., Lindqvist, Y., and Smith, S. (1999). Conversion of a β-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry*, 38(36):11643–11650.

Wolf, H., Assmann, D., and Fischer, E. (1978). Pulvomycin, an inhibitor of protein biosynthesis preventing ternary complex formation between elongation factor Tu, GTP, and aminoacyl-tRNA. *Proceedings of the National Academy of Sciences*, 75(11):5324–5328.

World Health Organization (2014). WHO fact sheet on antimicrobial resistance (AMR) N°194.

Wright, G. D. and Thompson, P. R. (1999). Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in bioscience : a journal and virtual library*, 4:D9–21.

Wu, Y., Kang, Q., Shen, Y., Su, W., and Bai, L. (2011). Cloning and functional analysis of the naphthomycin biosynthetic gene cluster in *Streptomyces* sp. cs. *Mol Biosyst*, 7(8):2459–2469.

Xiong, Y., Wu, X., and Mahmud, T. (2005). A homologue of the *Mycobacterium tuberculosis* PapA5 protein, Rif-Orf20, is an acetyltransferase involved in the biosynthesis of antitubercular drug rifamycin B by *Amycolatopsis mediterranei* S699. *Chembiochem*, 6(5):834–837.

Xu, J., Mahmud, T., and Floss, H. G. (2003). Isolation and characterization of 27-*O*-demethylrifamycin SV methyltransferase provides new insights into the post-PKS modification steps during the biosynthesis of the antitubercular drug rifamycin B by *Amycolatopsis mediterranei* S699. *Archives of biochemistry and biophysics*, 411(2):277–288.

Xu, J., Wan, E., Kim, C.-J., Floss, H. G., and Mahmud, T. (2005). Identification of tailoring genes involved in the modification of the polyketide backbone of rifamycin B by *Amycolatopsis mediterranei* S699. *Microbiology*, 151(Pt 8):2515–2528.

Yanai, K. and Murakami, T. (2004). The kanamycin biosynthetic gene cluster from *Streptomyces kanamyceticus*. *The Journal of Antibiotics*, 57(5):351–354.

Yanai, K., Murakami, T., and Bibb, M. (2006). Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*. *Proceedings of the National Academy of Sciences*, 103(25):9661–9666.

Yoneyama, H. and Katsumata, R. (2006). Antibiotic resistance in bacteria and its future for novel antibiotic development. *Bioscience, biotechnology, and biochemistry*, 70(5):1060–1075.

You, Y.-O., Khosla, C., and Cane, D. E. (2013). Stereochemistry of reductions catalyzed by methyl-epimerizing ketoreductase domains of polyketide synthases. *Journal of the American Chemical Society*, 135(20):7406–7409.

Yu, J. and Tao, M. (2010). Effect of inorganic salts on the conjugation and heterologous expression of actinorhodin in *Streptomyces avermitilis*. *Wei sheng wu xue bao = Acta microbiologica Sinica*, 50(11):1556–1561.

Yu, T. W., Müller, R., Muller, M., Zhang, X., Draeger, G., Kim, C. G., Leistner, E., and Floss, H. G. (2001). Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. *Journal of Biological Chemistry*, 276(16):12546–12555.

Yu, T.-W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B. S., Hutchinson, C. R., and Floss, H. G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proceedings of the National Academy of Sciences*, 96(16):9051–9056.

Yuan, H., Zhao, W., Zhong, Y., and Wang, J. (2011). Two genes, *rif15* and *rif16*, of the rifamycin biosynthetic gene cluster in *Amycolatopsis mediterranei* likely encode a transketolase and a P450 monooxygenase, respectively, both essential for the conversion of rifamycin SV into B. *Acta Biochimica et Biophysica Sinica*.

Zabala, D., Braña, A. F., Flórez, A. B., Salas, J. A., and Méndez, C. (2013). Engineering precursor metabolite pools for increasing production of antitumor mithramycins in *Streptomyces argillaceus*. *Metabolic Engineering*, 20(C):187–197.

Zerbe, K., Pylypenko, O., Vitali, F., Zhang, W., Rouset, S., Heck, M., Vrijbloed, J. W., Bischoff, D., Bister, B., Süssmuth, R. D., Pelzer, S., Wohlleben, W., Robinson, J. A., and Schlichting, I. (2002). Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *Journal of Biological Chemistry*, 277(49):47476–47485.

Zweerink, M. M. and Edison, A. (1987). Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by *Bacillus subtilis*. III. Mode of action of difficidin. *The Journal of Antibiotics*, 40(12):1692–1697.

Appendix A

List of strains

Table A.1: List of strains used and generated in this work.

Strains	Description	Reference or source
E. coli DH5α	Strain used for routine cloning. F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k-, m_k+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Grant et al., 1990
E. coli TOP10	Strain used for routine cloning. F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rpsL*(*Str*^{R}) *end*A1 λ-	
E. coli DH10B	Strain used for routine cloning. F- *end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*L Δ*lac*X74 Φ80*lac*ZΔM15 *ara*D139 Δ(*ara-leu*)7697 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) λ-	Grant et al., 1990
E. coli ET12567/ pUZ8002	Methylation deficient strain used for conjugation with *Streptomyces*. pUZ8002 provides conjugation machinery. *dam13::Tn9* (Chloramphenicol resistance) *dcm-6 hsdM hsdR recF143 zjj-201::Tn10 galK2 galT22 ara14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtlI glnV44*, pUZ8002 (kanamycin resistance)	MacNeil et al., 1992; pUZ8002, J. Wilson and D. Figurski, unpublished
E. coli TOP10/ pR9406	Strain used for routine cloning carrying conjugation plasmid pR9406 (carbenicillin resistance)	pR9406, A. Siddique and D. Figurski, unpublished
S. leeuwenhoekii C34	Wild type strain	Busarakam et al., 2014
S. coelicolor M1152	M145, Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB*[*C1298T*] ([S433L])	Gomez-Escribano and Bibb, 2011
S. coelicolor M1650	M1152 carrying pIJ12853	This work
S. leeuwenhoekii M1653	Δ*cxmK::neo*	This work

Continued. List of strains used and generated in this work.

Strains	Description	Reference or source
S. leeuwenhoekii M1654	M1653 carrying pIJ12852	This work
S. leeuwenhoekii M1655	*S. leeuwenhoekii* C34 carrying pGUS	This work
S. leeuwenhoekii M1656	*S. leeuwenhoekii* C34 carrying pIJ10740	This work
S. leeuwenhoekii M1660	*S. leeuwenhoekii* C34 carrying pSET152	This work
S. leeuwenhoekii M1661	*S. leeuwenhoekii* C34 carrying pGM1190	This work
Bacillus subtilis EC1524	Indicator microorganism for bioassays	Institute for Food Research, Norwich, UK
Micrococcus luteus ATCC 4698	Indicator microorganism for bioassays	ATCC

Appendix B

List of plasmids

Table B.1: List of plasmids used and generated in this work.

Plasmid	Description	Reference or source
pBluescript II SK(+)	General cloning vector, *ampR* (ampicillin resistance)	Alting-Mees and Short, 1989
pKC1132	Cloning vector, suicide, conjugative (*oriT* from RK2), *aac(3)IV* (apramycin resistance)	Bierman et al., 1992
pSET152	Cloning vector, conjugative (*oriT* from RK2), *aac(3)IV* (apramycin resistance), integrative (ΦC31 *attP*)	Bierman et al., 1992
pTC192-Km	Source of *neo* (kanamycin resistance gene)	Rodriguez-Garcia et al., 2006
pGM1190	pSG5 derivative, self-replicative, temperature sensitive, *tsr*, *aac(3)IV* (apramycin resistance), *oriT*, *to* terminator, P_{tipA}, RBS, *fd* terminator	Muth et al., 1989; pGM1190, G. Muth, unpublished
pESAC13	PAC vector (P1-phage replicon) for genomic library construction; conjugative (*oriT* from RK2), integrative (ΦC31 *attP*), *tsr* (thiostrepton resistance in *Streptomyces*), *neo* (kanamycin resistance in *E. coli*), *P1 rep*, *sacBII*	Sosio et al., 2000; pESAC13, M. Sosio, unpublished
pGUS	Promoterless *gusA* (β-glucuronidase gene, codon-optimised for streptomycetes) in pSET152. Used as a promoter detection vector	Myronovskyi et al., 2011
pIJ10740	pGUS derivative with *ermE** promoter driving *gusA* transcription	Morgan Feeney, unpublished
pIJ10257	Expression vector for *Streptomyces*, with *ermE** promoter, *hyg* (hygromycin B resistance), conjugative (*oriT* from RK2), integrative (ΦBT1 *attP*)	Hong et al., 2005
pIJ12850	Derivative of pGM1190 with JFC010/JFC009-*neo*-JFC011/JFC012 fragments; for deletion of AHBA synthase (*cxmK*) in *S. leeuwenhoekii* C34	This work

Continued. List of plasmids used and generated in this work.

Plasmid	Description	Reference or source
pIJ12851	Derivative of pKC1132 with JFC010/JFC009-*neo*-JFC011/JFC012 fragments; for deletion of *cxmK* in *S. leeuwenhoekii* C34	This work
pIJ12857	Derivative of pKC1132 with JFC010/JFC009 fragment; for deletion of *cxmK* in *S. leeuwenhoekii* C34	This work
pIJ12858	Derivative of pKC1132 with JFC011/JFC012 fragment; for deletion of *cxmK* in *S. leeuwenhoekii* C34	This work
pIJ12859	Derivative of pKC1132 with JFC010/JFC009-JFC011/JFC012 fragment for the deletion of *cxmK* in *S. leeuwenhoekii* C34	This work
pIJ12860	Derivative of pGM1190 with JFC010/JFC009-JFC011/JFC012 fragment for the deletion of *cxmK* in *S. leeuwenhoekii* C34	This work
pIJ12861	pBluescript II SK(+) derivative with JFC026/JFC034 fragment for cloning of *cxmK*	This work
pIJ12852	pIJ10257 derivative with *cxmK* for complementation of Δ*cxmK::neo* mutant (M1653)	This work
pIJ12853	Derivative of pESAC13; contains 145 kb of *S. leeuwenhoekii* C34 chromosome including the chaxamycin biosynthesis gene cluster	This work

Appendix C

List of primers

Table C.1: List of primers used in this work.

Name	Sequence 5' → 3'	Notes
LF044F	cgaagatcccgtcgatgatgt	Amplify a region of 118 bp that results after the integration of pSET152 at the ΦC31 *attB* site of *S. leeuwenhoekii* C34 genome (Foulston and Bibb, 2010)
LF045R	cacaacccttgtgtcatgtc	
JFC009	tttctagagaagccatcagccaggact	Amplify upstream flanking region of *cxmK*; contain restriction sites for *Xba*I and *Hind*III, respectively
JFC010	ttaagcttgtacgtccagcagctccagt	
JFC011	tttctagagctcggtctcgtcgtactg	Amplify downstream flanking region of *cxmK*; contain restriction sites for *Xba*I and *Bam*HI, respectively
JFC012	ttggatccactgctcctgctcgactacg	
JFC022	cgattcgaagagtgacagca	Screening of genomic library for chaxamycin gene cluster
JFC023	gttgtaggtggcgattttgc	
JFC024	gcgaagagaagagcgaggt	Screening of genomic library for chaxamycin gene cluster
JFC025	cacacattgctcagatgcac	
JFC026	agacatatgaacgcgcgacag	Cloning of *cxmK* in pIJ10257; contain restriction sites for *Nde*I and *Hind*III, respectively
JFC034	aagcttacgacctggtag	
JFC032	gtcgttccctcatttcctca	For confirmation of deletion of AHBA synthase gene
JFC033	gacctggtagccggtgtg	
JFC036	gacgagaacaccacgaaggt	To resolve a frame-shift within one of the PKS genes
JFC037	gaccaccggcactactgg	
JFC053	aagcaggtcagcaggatca	To verify the sequence between *cxm24* and *cxmY*
JFC054	ttccggaagatgaacgtgac	
27F	agagtttgatcatggctcag	Amplification of 16S rRNA gene
1492R	tacggttaccttgttacgactt	

cxmK is AHBA synthase gene of *S. leeuwenhoekii* C34.

Appendix D

Maps of vectors and constructions

GenBank files of genetic constructions and vectors were generated manually in ApE software version 2.0.46 (http://biologylabs.utah.edu/jorgensen/wayned/ape/); drawings and manual annotation of plasmids was done in SnapGene Viewer software (http://www.snapgene.com); final edition of the Figures was carried out in Inkscape software version 0.91 (https://inkscape.org/en/).

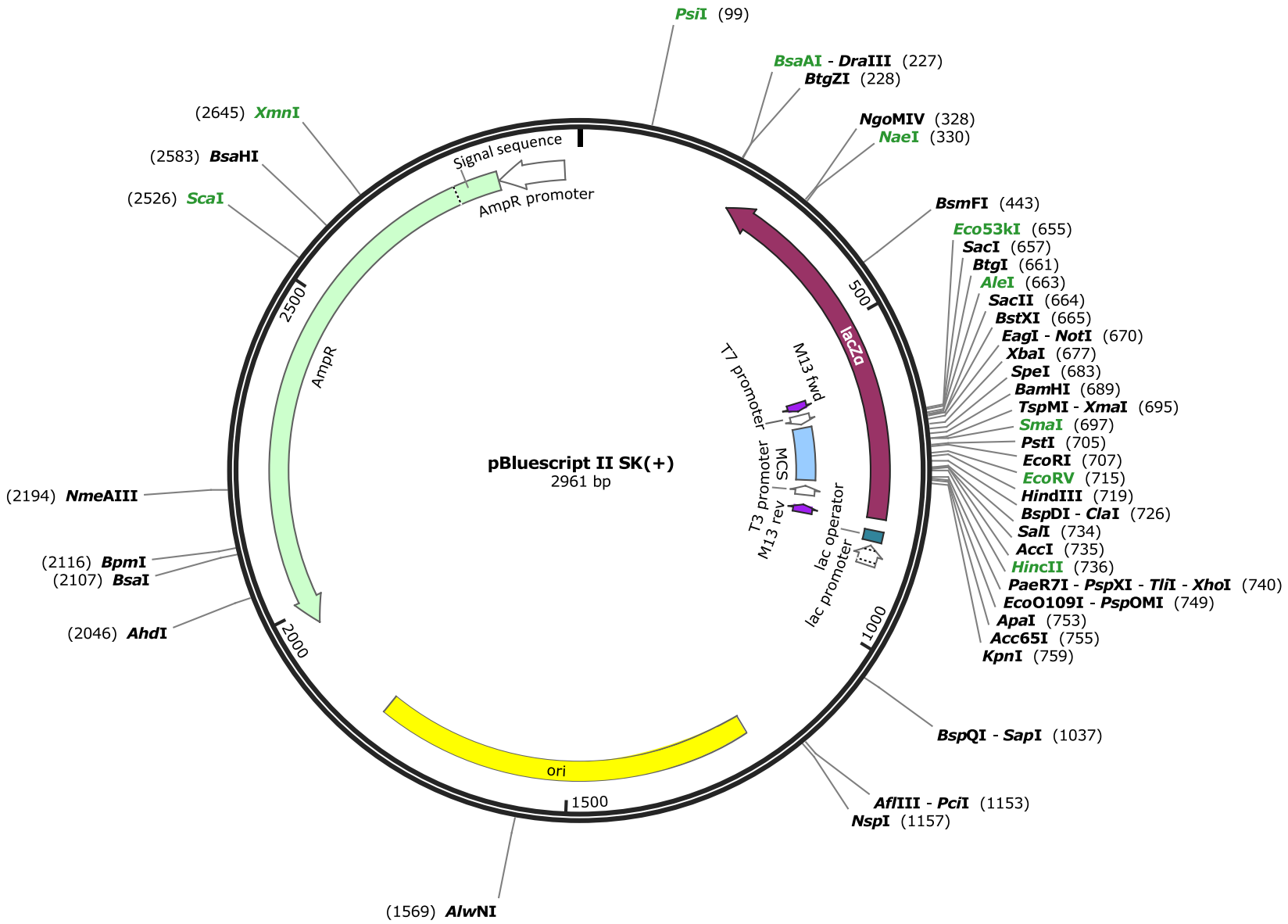


Figure D.1: Map of pBluescript II SK(+). Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

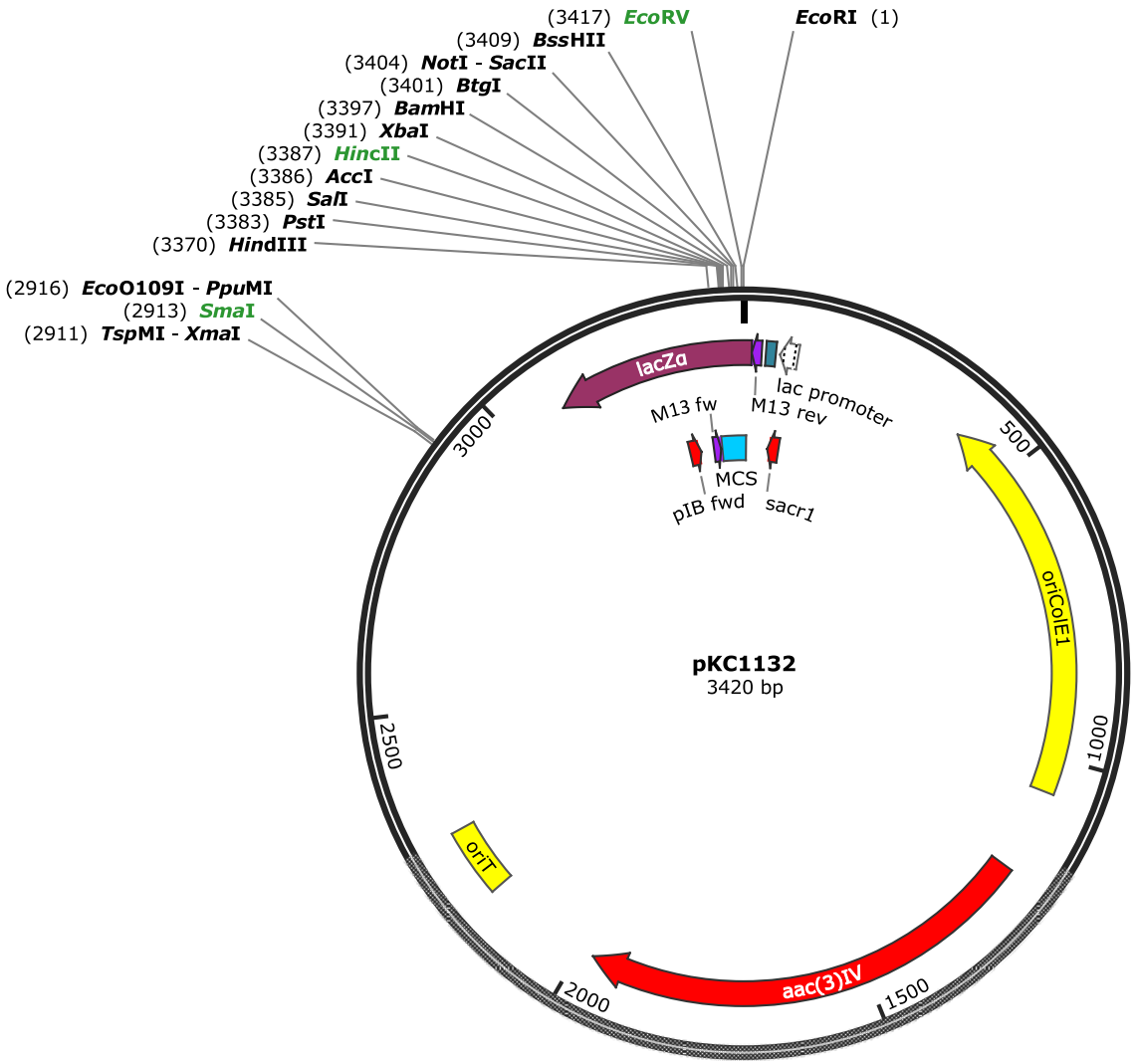


Figure D.2: Map of pKC1132. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

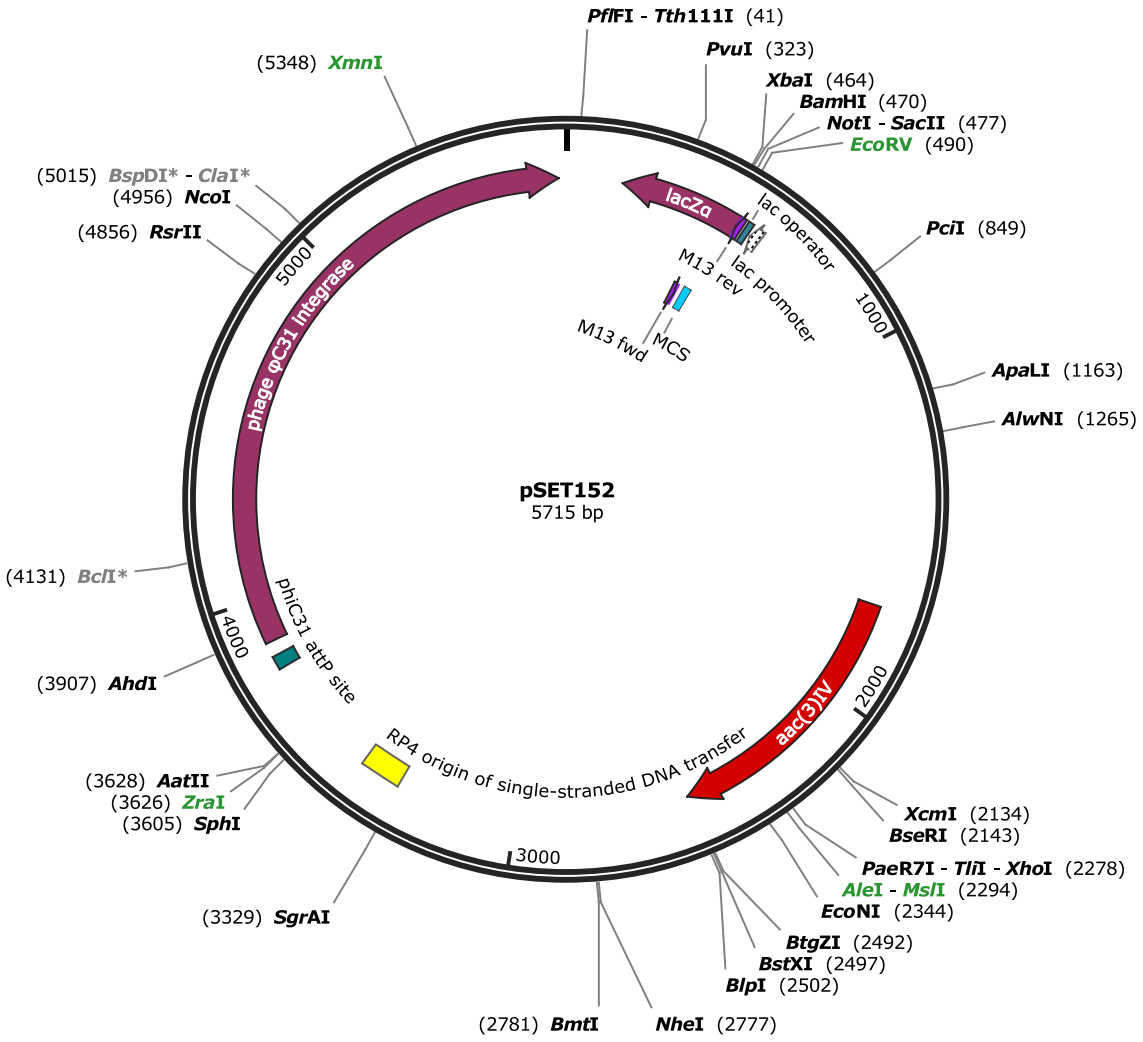


Figure D.3: Map of pSET152. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

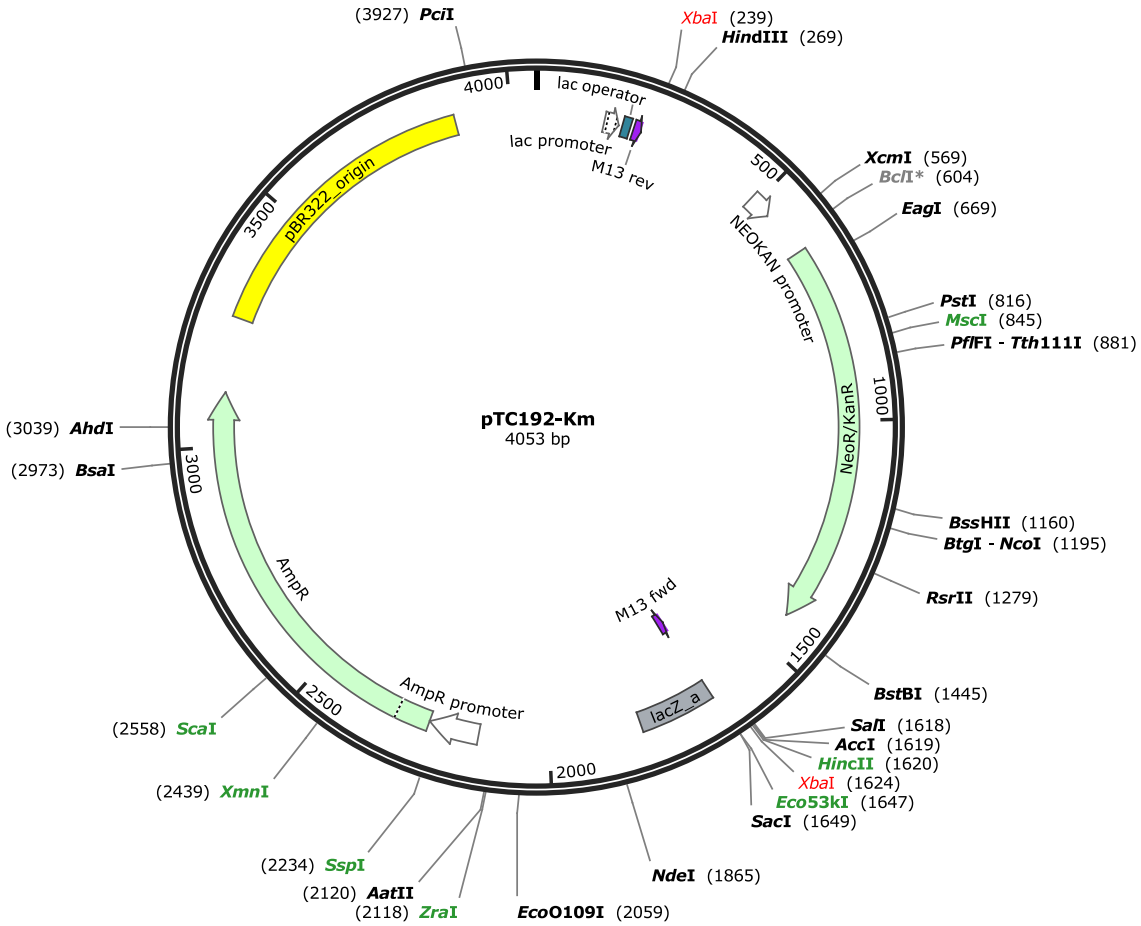


Figure D.4: Map of pTC192-Km. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green; *Xba*I sites are highlighted in red to show the region in which the *neo* cassette is located.

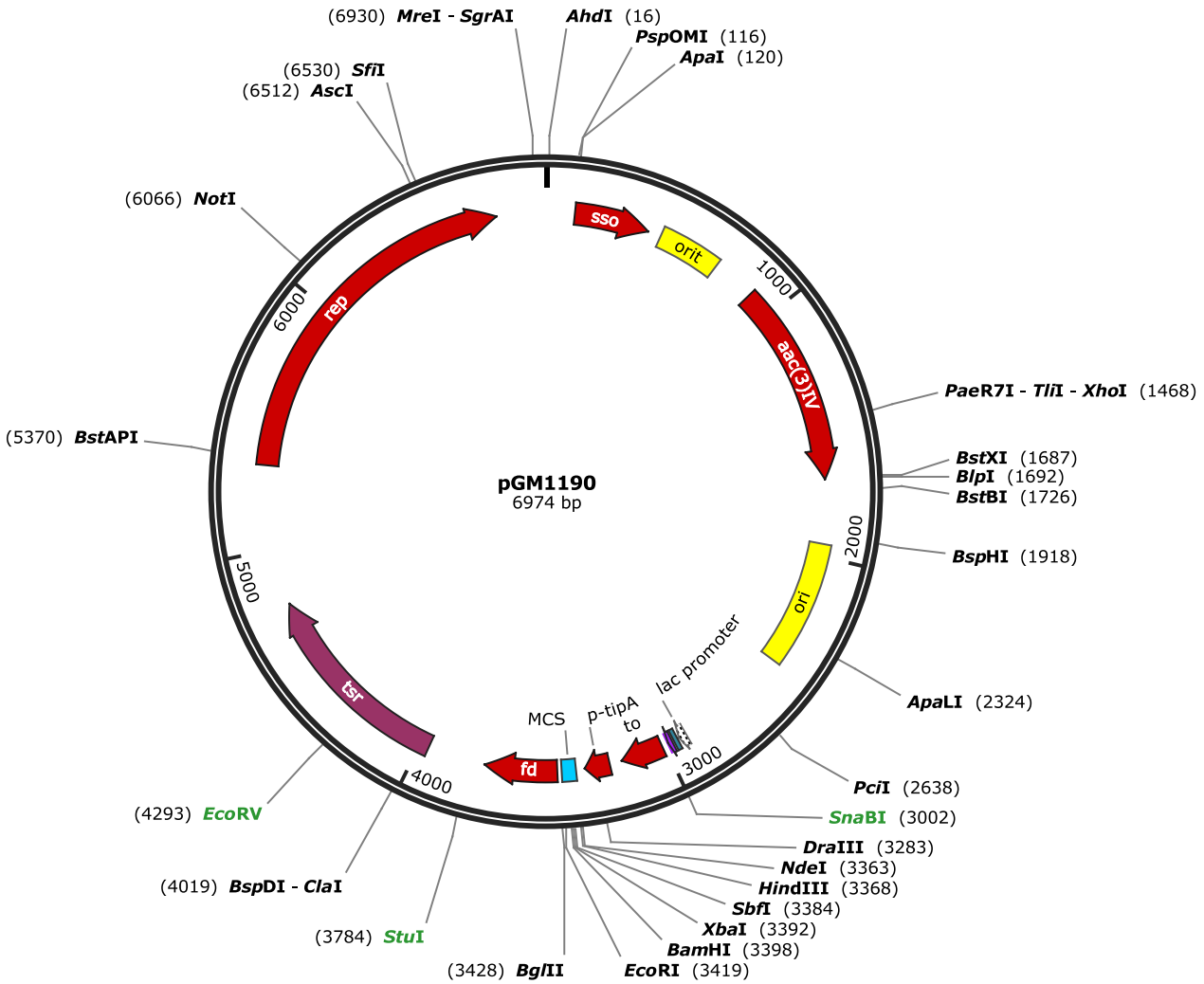


Figure D.5: Map of pGM1190. High-copy temperature-sensitive vector in streptomycetes, used for deletion of genomic regions in *S. leeuwenhoekii* C34. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

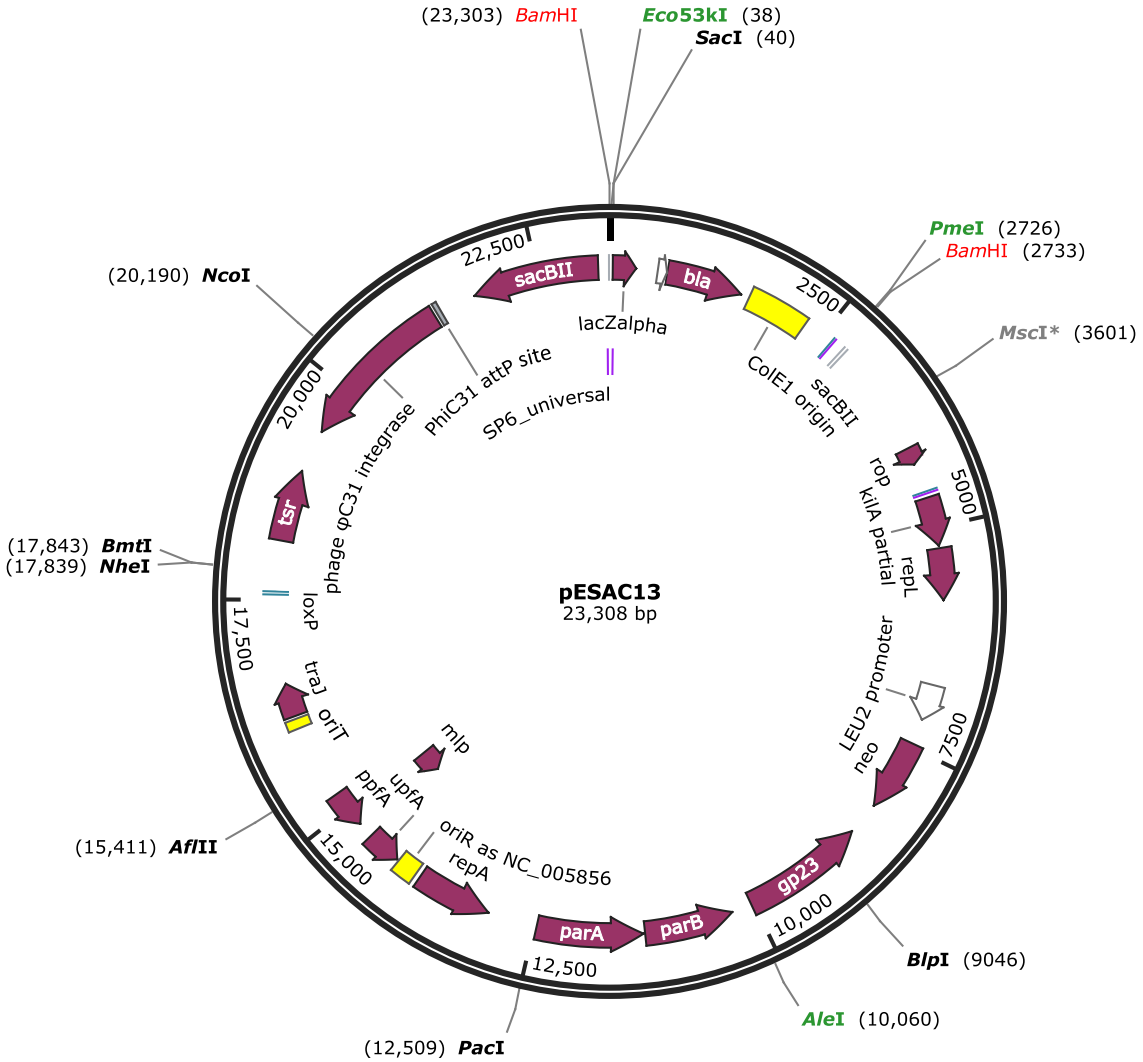


Figure D.6: Map of pESAC13. Vector used for cloning a genome library of *S. leeuwenhoekii* C34; NCBI accession: LM999999.1. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green; *Bam*HI sites are highlighted in red to precise the site in which the genome library *S. leeuwenhoekii* C34 was cloned. Segment from nt position: 22 to 2707 comes from pUC19; 4659 to 6245 and 8667 to 15349 come from phage P1; 6246 to 7331 has 100% identity with pCYPAC6 (accession AF133437); 7332 to 8566 comes from TN903; 19436 to 21527 comes from pSET152.

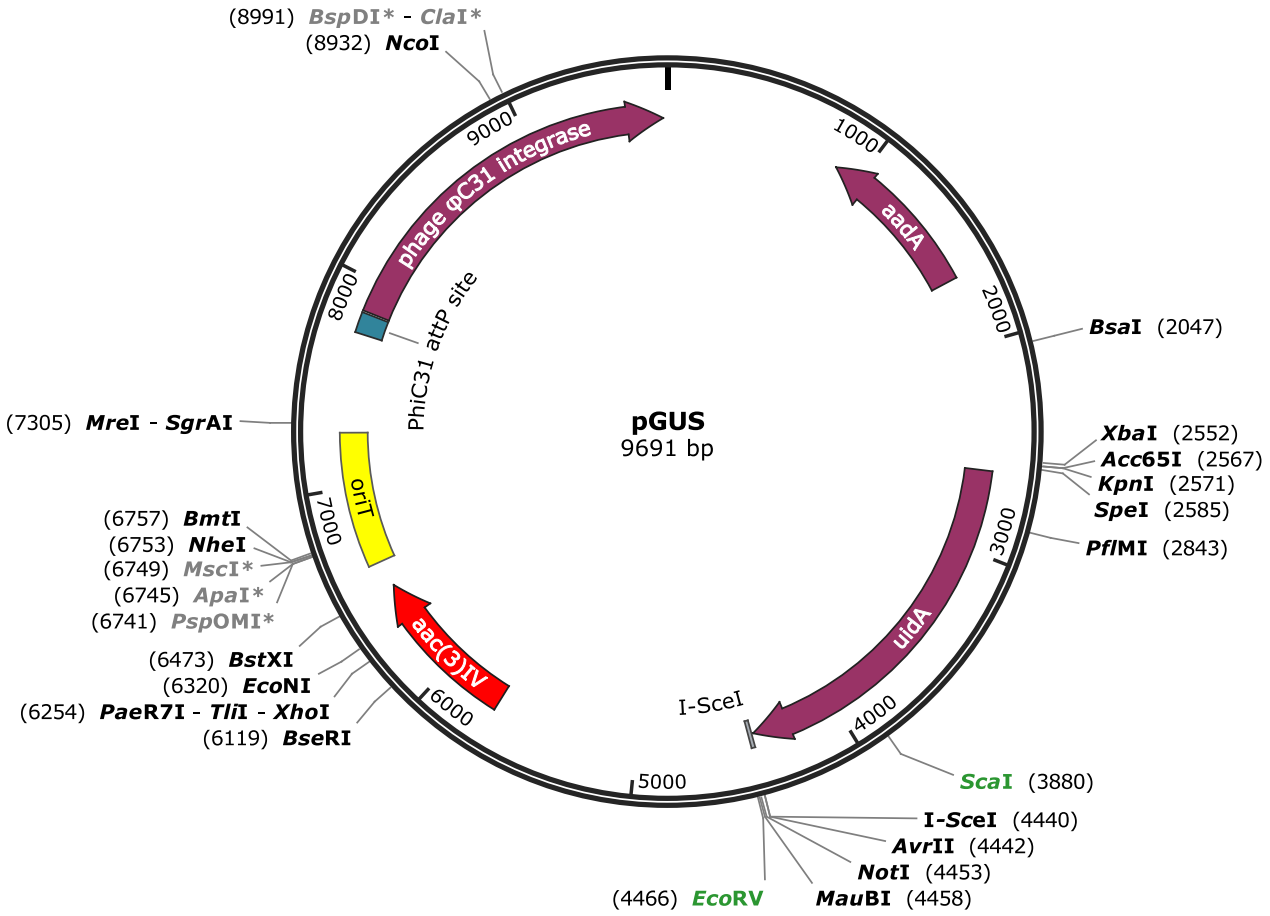


Figure D.7: Map of pGUS. pGUS is a derivative of pSET152 that carries the β-glucuronidase gene that was codon-optimised for streptomycetes (*uidA*) (Myronovskyi et al., 2011). Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green. The GenBank file of the sequence of this vector was kindly provided by Dr. Maksym Myronovskyi.

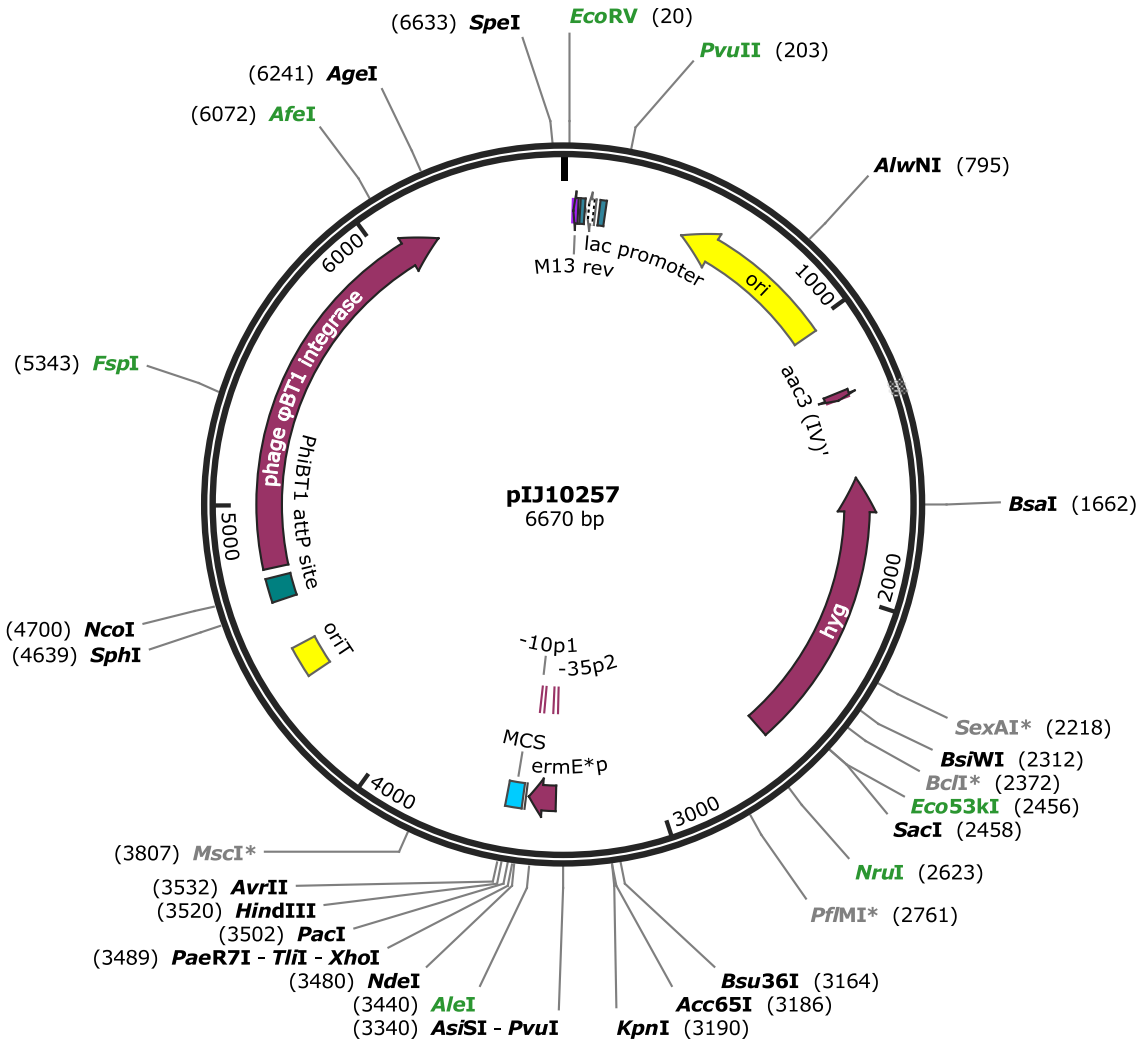


Figure D.8: Map of pIJ10257. Used for cloning of genes under the constitutive promoter *ermE**. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

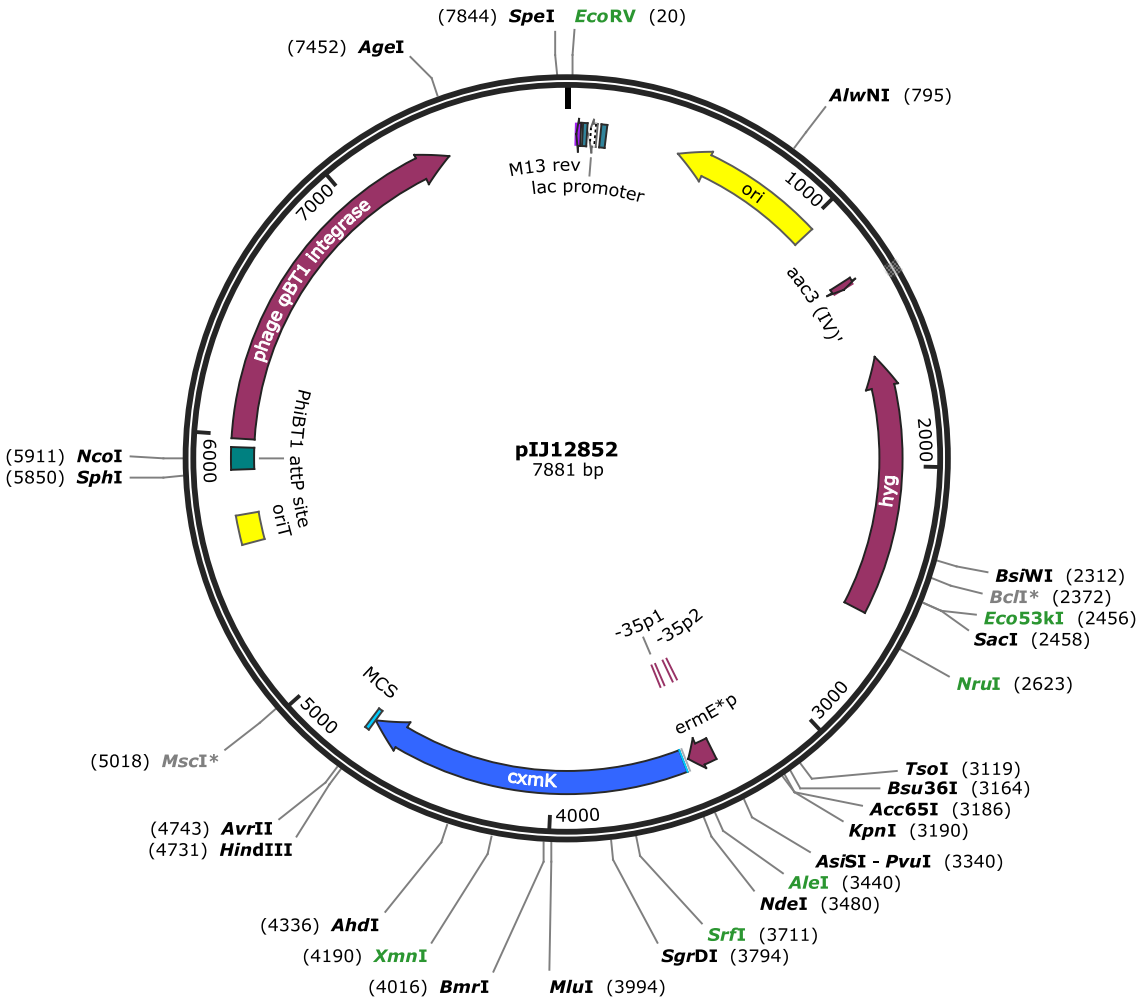


Figure D.9: Map of pIJ12852. A derivative of pIJ10257 that carries the AHBA synthase gene (*cxmK*) of *S. leeuwenhoekii* C34, whose transcription is driven by the constitutive promoter *ermE**. The fragment was amplified with primers JFC026/JFC034. Set of enzymes that recognise unique cleavage sites are shown; blunt cutters are highlighted in green.

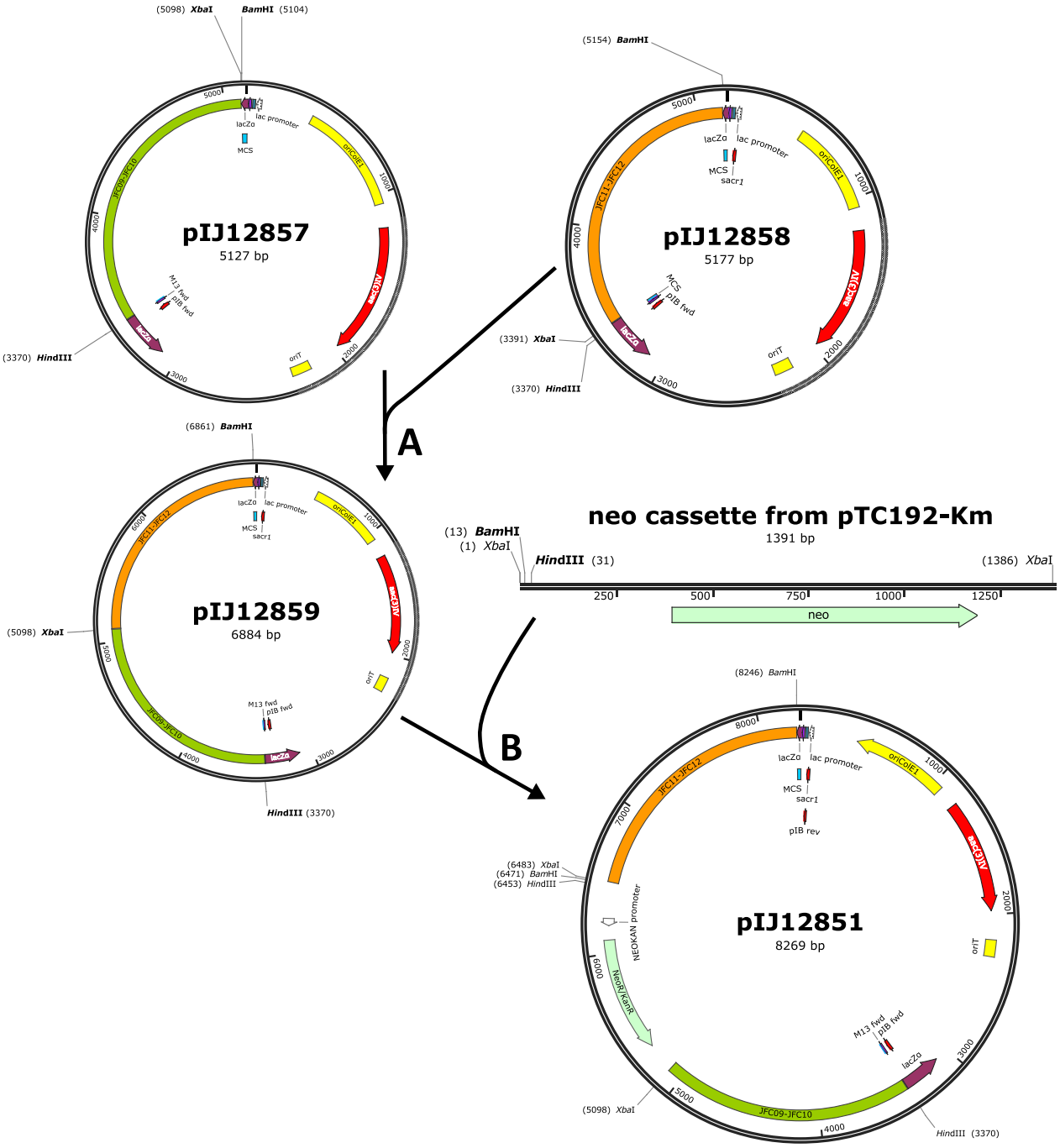


Figure D.10: Construction of pIJ12851. pIJ12851 is a derivative of pKC1132 (Bierman et al., 1992) that was initially constructed for the deletion of AHBA synthase gene (*cxmK*) in *S. leeuwenhoekii* C34. Two DNA regions of about 1.7 kb that flank *cxmK* were PCR-amplified with primers JFC010/JFC009 (containing 5' *Hind*III and *Xba*I sites, respectively) and JFC011/JFC012 (containing 5' *Xba*I and *Bam*HI sites, respectively), then these fragments were separately cloned into pKC1132, previously digested with *Hind*III plus *Xba*I and *Bam*HI plus *Xba*I to yield pIJ12857 and pIJ12858, respectively. (**A**) pIJ12859 results from cloning the 1.7 kb fragment contained in pIJ12858 into pIJ12857. (**B**) Finally, pIJ12851 results from cloning the *neo* cassette (1.3 kb in length) obtained from the digestion of pTC192-Km with *Xba*I; the 1.3 kb band was purified from a 1% agarose gel and cloned into pIJ12859, previously digested with *Xba*I. Relevant restriction enzyme sites are shown.

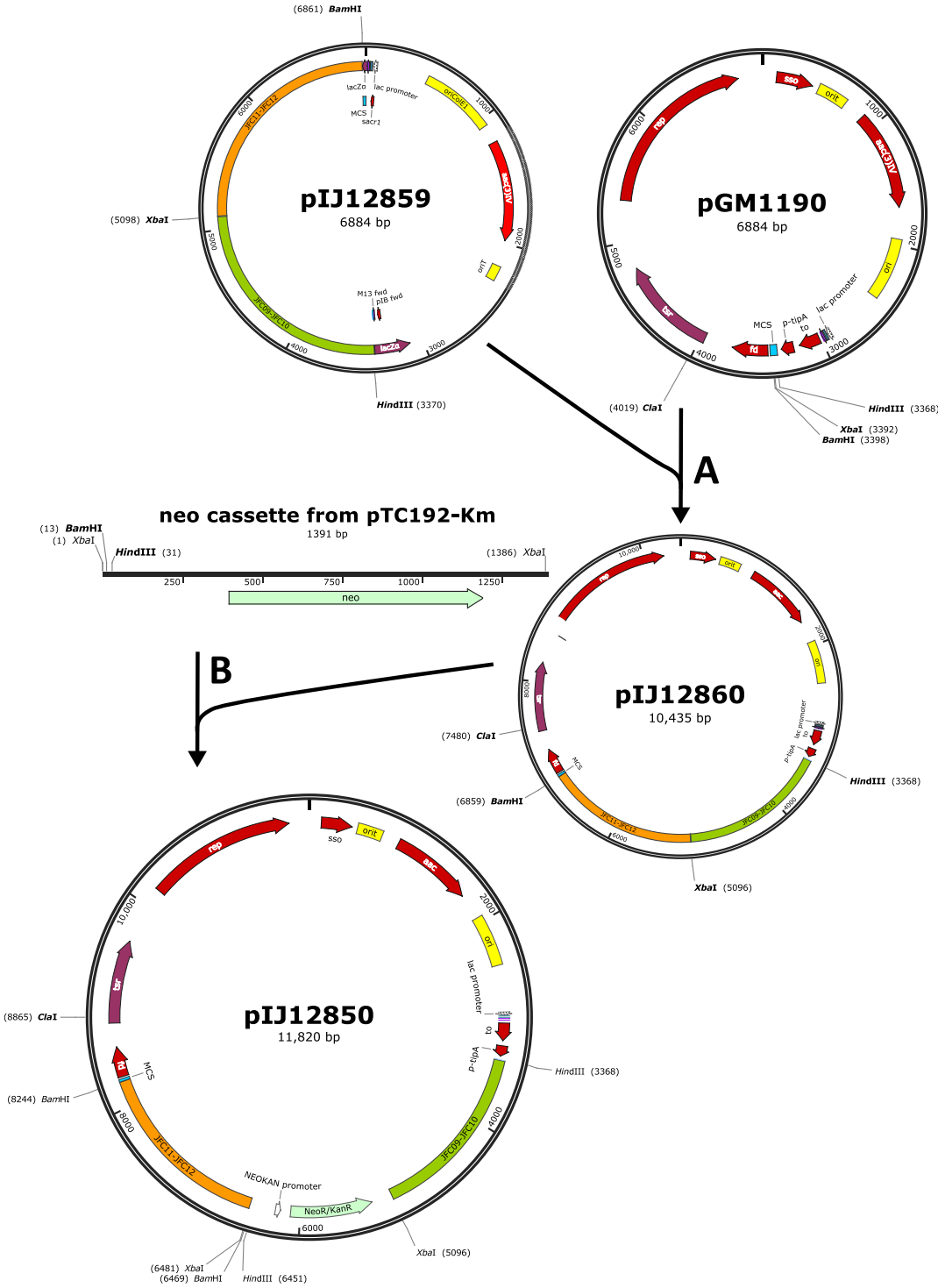


Figure D.11: Construction of pIJ12850. pIJ12850 is a derivative of pGM1190 (Muth et al., 1989), constructed for the deletion of AHBA synthase gene (*cxmK*) in *S. leeuwenhoekii* C34, using double-crossover recombination to replace *cxmK* with the kanamycin resistance gene, *neo*. Two DNA regions that flank *cxmK* were previously cloned into pIJ12859 as JFC09/JFC10-JFC11/JFC12 and excised from it by digestion with *Hind*III, *Bam*HI and *Cla*I, the later enzyme was used to allow the purification of the fragment JFC09/JFC10-JFC11/JFC12 (3.4 kb), which was similar in length to the backbone vector (about 3.4 kb). (**A**) The fragment JFC09/JFC10-JFC11/JFC12 was purified from a 1% agarose gel and cloned into pGM1190, previously digested with *Hind*III, *Bam*HI, to yield pIJ12860. (**B**) pIJ12850 results from cloning the *neo* cassette (1.3 kb length) obtained from the digestion of pTC192-Km with *Xba*I; the 1.3 kb band was purified from a 1% agarose gel and cloned into pIJ12860, previously digested with *Xba*I. Relevant restriction enzyme sites are shown.

Appendix E

Media, solutions and buffers recipes

All formulation recipes for media, solutions and buffers preparation are listed in alphabetic order.

E.1 Agar media recipes

DNA medium contains per 200 ml distilled water (Kieser et al., 2000):

Difco Nutrient Agar	4.6 g

ISP2 agar medium contains per litre of distilled water (Shirling and Gottlieb, 1966):

Yeast extract	4 g
Malt extract	10 g
Glucose	4 g
Agar	20 g

LA medium contains per litre of distilled water (Sambrook et al., 1989):

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g

Dissolve the ingredients, except agar, in the distilled water and pour 200 ml into 250 ml Erlenmeyer flasks each containing 2 g agar. Close the flasks and autoclave.

SFM agar medium contains per litre of tap water (Hobbs et al., 1989):

Agar	20 g
Mannitol	20 g
Soya flour	20 g

Soya flour was purchased from the local supermarket. The brand used in this work was *Mi tierra* from Nutrisa - Nutrición y Alimentos S. A., Santiago, Chile. Its composition per 100 g is listed as follows:

Proteins	50.0 g
Fats	1.0 g
Carbohydrates	30.0 g
Sodium	0.010 g
Magnesium	0.300 g
Phosphorus	0.730 g
Potassium	2.470 g
Iron	0.010 g
Copper	0.001 g
Calcium	0.320 g
Zinc	0.005 g

R2-S medium based on the original recipe (Hopwood and Wright, 1978; Okanishi et al., 1974). The recipe to prepare R2 medium (Kieser et al., 2000) has been modified and does not include sucrose. Prepare 800 ml of distilled water with the following components:

K_2SO_4	0.25 g
$MgCl_2 \cdot 6H_2O$	10.12 g
Glucose	10 g
Casaminoacids	0.1 g

In an Erlenmeyer flasks add 2.2 g of agar plus 80 ml of the solution and autoclave. At the time of using, melt the medium and add the following autoclaved solutions in the order listed:

KH_2PO_4 (0.5%)	1 ml
$CaCl_2 \cdot 2H_2O$ (3.68%)	8 ml
L-proline (20%)	1.5 ml
TES buffer (5.73%)	10 ml
Trace element solution	0.2 ml
NaOH (1N)	0.5 ml

TES buffer (5.73%) was previously adjusted to pH 7.2; it is not strictly necessary to use sterile NaOH 1N solution; the recipe to prepare trace elements solution is in the solutions section (Appendix E.4).

E.2 Liquid media recipes

2xYT medium contains per litre of distilled water (Kieser et al., 2000):

Tryptone	16 g
Yeast Extract	10 g
NaCl	5 g

ISP2 medium contains per litre of distilled water (Shirling and Gottlieb, 1966):

Yeast extract	4 g
Malt extract	10 g
Glucose	4 g

LB medium contains per litre of distilled water (Sambrook et al., 1989):

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Modified ISP2 medium contains per litre of distilled water (Rateb et al., 2011a):

Yeast extract	4 g
Malt extract	10 g
Glycerol	10 g

Adjust to pH 7.2 by adding NaOH.

R3 medium contains per litre of distilled water (Shima et al., 1996):

Glucose	10 g
Yeast Extract	5.0 g
Casaminoacids	0.1 g
L-proline	3 g
$MgCl_2 \cdot 6H_2O$	10 g
$CaCl_2 \cdot 2H_2O$	4 g
K_2SO_4	0.2 g
KH_2PO_4	0.05 g
TES	5.6 g

Supplement the medium with 0.2 ml of trace elements solution per 200 ml of solution (Appendix E.4). Adjust to pH 7.2 by adding NaOH.

SOB and SOC media contains per litre of distilled water (Hanahan, 1983):

Tryptone	20 g
Yeast extract	5 g
NaCl (10 mM)	0.584 g
KCl (2.5 mM)	0.186 g
$MgCl_2 \cdot 6H_2O$ (10 mM)	2.033 g
$MgSO_4 \cdot 7H_2O$ (10 mM)	2.467 g
Glucose (20 mM) **for SOC only**	3.603 g

Adjust to pH 7.0 by adding NaOH.

TSB medium contains per litre of distilled water (Kieser et al., 2000):

Tryptone soya broth	30 g

YEME medium contains per litre of distilled water (Kieser et al., 2000):

Yeast extract	3 g
Peptone	5 g
Malt extract	3 g
Glucose	10 g
Sucrose	340 g

After autoclaving add 2 ml of autoclaved 2.5 M $MgCl_2$ solution.

E.3 Solution recipes for microbiology

Solutions used for microbiology were prepared as follows: register weight of an empty 15 ml conical centrifuge tube (X_1 g); add an amount of the chemical into the tube and register weight (X_2 g), do this inside a fume hood since some chemical powder are toxic. Calculate the exact weight of the chemical previously added by subtracting $X_2 - X_1$, and add the volume of solvent needed to reach the desired concentration. Finally, filter the solution through a 0.22 µm filter and aliquot in sterile 1.5 ml Eppendorf tubes. Store at -20 °C. Use a water bath set at 37 °C for thawing.

E.3.1 Antibiotic solutions

Apramycin solution. Apramycin sulphate Sigma cat. no. A2024 was dissolved in distilled water at a final concentration of 50 mg/ml.

Carbenicillin solution. Carbenicillin disodium salt Invitrogen cat. no. 10177-012 was dissolved in distilled water at a final concentration of 50 mg/ml.

Chloramphenicol solution. Chloramphenicol Calbiochem cat. no. 220551 was dissolved in absolute ethanol at a final concentration of 50 mg/ml.

Kanamycin solution. Kanamycin Gibco cat. no. 11815-024 was dissolved in distilled water at a final concentration of 50 mg/ml. It is recommended to use it with low concentrations of salts; for example, 60 mM $CaCl_2$ or 60 mM $MgCl_2$ have been used without compromising antibiotic effectiveness.

Hygromycin B solution. Hygromycin B from *Streptomyces hygroscopicus* was purchased from Sigma cat. no. 238813 in phosphate buffer solution (PBS) at a concentration of 50 mg/ml. The solution was aliquoted in 1.5 ml Eppendorf tubes and wrapped in aluminium foil to protect the solution from the light (hygromycin B is light-sensitive). Use in agar or liquid culture solutions with low salt concentrations; for example 10 mM $CaCl_2$ or 10 mM $MgCl_2$ were used without compromising antibiotic effectiveness, but 60 mM $CaCl_2$ or 60 mM $MgCl_2$ must not be used.

Nalidixic acid solution. Nalidixic acid Sigma cat. no. N8878 was dissolved in 0.3 M NaOH solution at a final concentration of 25 mg/ml.

Thiostrepton solution. Thiostrepton Sigma cat. no. T8902 was dissolved in DMSO solution at a final concentration of 50 mg/ml.

E.3.2 Other solutions

IPTG. IPTG Fermentas cat. no. R0392 was dissolved in distilled water at a final concentration of 0.1 M. Molecular weight of IPTG is 238.3 g/mol.

Lysozyme solution. Lysozyme from chicken egg Sigma cat. no. L6876 was dissolved in distilled water at a final concentration of 50 mg/ml.

X-Gal solution. X-Gal Fermentas cat. no. R0404 was dissolved in DMSO solution at a final concentration of 20 mg/ml. Eppendorf tubes were wrapped with aluminium foil to protect from light.

E.4 Buffer and solution recipes

List of solution stocks used to prepare buffers and other solutions.

EDTA ..... 0.25 M pH 8
NaCl ..... 1 M
NaOH ..... 1 N
Tris-HCl ..... 2 M pH 8
SDS (from supplier) 20%

Compound (molecular weight): EDTA (372.24 g/mol); NaCl (58.44 g/mol); potassium acetate (98.14 g/mol); SDS (288,37 g/mol); TRIS (121.14 g/mol).

N3 buffer contains 3 M potassium acetate (or sodium acetate). Adjust to pH 5.5 with acetic acid.

Normal saline solution contains 0.9% w/v NaCl in distilled water.

P1 buffer contains per 50 ml of distilled water:

EDTA (stock solution)	2 ml
Tris-HCl (stock solution)	1.25 ml
Water	46.75 ml

The final concentration of each component is: 10 mM EDTA; 50 mM Tris-HCl.

P2 buffer contains per 50 ml of distilled water:

NaOH (stock solution)	10 ml
SDS (stock solution)	2.5 ml
Water	37.5 ml

The final concentration of each component is: 200 mM NaOH; 1% SDS.

SET buffer contains per 50 ml of distilled water:

NaCl (stock solution)	3.75 ml
EDTA (stock solution)	5 ml
Tris-HCl (stock solution)	0.50 ml
Water	40.75 ml

The final concentration of each component is: 75 mM NaCl; 25 mM EDTA; 20 mM Tris-HCl.

STET buffer contains per 50 ml of distilled water:

Sucrose (8% w/v)	4 g
EDTA (stock solution)	10 ml
Tris-HCl (stock solution)	1.25 ml
Triton X-100 (5% v/v)	2.5 ml

The final concentration of each component is: 50 mM EDTA; 50 mM Tris-HCl.

Trace elements solution contains per litre of distilled water (Kieser et al., 2000):

$ZnCl_2$	40 mg
$FeCl_3 \cdot 6\,H_2O$	200 mg
$CuCl_2 \cdot 2\,H_2O$	10 mg
$MnCl_2 \cdot 4\,H_2O$	10 mg
$Na_2B_4O_7 \cdot 10\,H_2O$	10 mg
$(NH_4)_6Mo_7O_{24} \cdot 4\,H_2O$	10 mg

Store at 4 °C.

Tris-EDTA buffer contains per litre of distilled water (Kieser et al., 2000):

EDTA (stock solution)	4 ml
Tris-HCl (stock solution)	5 ml

The final concentration of each component is: 1 mM EDTA; 10 mM Tris-HCl.

Appendix F

Protocols used in this work

Media, solutions and buffer recipes are detailed in Appendix E.

F.1 Extraction of plasmid, PAC and genomic DNA

F.1.1 Boiling extraction of plasmid DNA (on small scale)

1. Set overnight cultures of *E. coli* strains in 1.5 ml Eppendorf tubes with 1 ml LB, supplemented with relevant antibiotics.
2. Centrifuge at 10,000 rpm for 5 min and discard supernatant.
3. Resuspend cell pellet in 200 µl STET buffer.
4. Add 10 µl 25 mg/ml lysozyme inside the cap of the Eppendorf tube and vortex the mixture for 1 min at room temperature.
5. Place the samples in boiling water for 45 sec.
6. Centrifuge at 10,000 rpm for 5 min and remove cell debris with a toothpick from the bottom of the Eppendorf tube.
7. Add 250 µl isopropanol and carefully mix by inversion. Let stand for 15 min at room temperature.
8. Centrifuge at 20,000 rpm for 20 min and discard supernatant.
9. Open the tubes and let them dry for 30 min.
10. Add 50 µl distilled H_2O and gently shake for 15 min. Keep at 4 °C.

F.1.2 PAC DNA extraction protocol

1. Set up an overnight culture of *E. coli* DH10B harbouring PAC DNA in 10 ml LB with relevant antibiotics at 37 °C and 250 rpm.
2. Centrifuge culture at 4,000 rpm for 5 min at room temperature.

3. Discard supernatant and resuspend cell pellet in 300 µl of P1 solution. Transfer the resuspended cell pellet into a 2 ml Eppendorf tube.
4. Add 350 µl of P2 solution and incubate for 2 min at room temperature.
5. Neutralise with 450 µl of N3 solution and incubate on ice for 10 min.
6. Centrifuge at 13,000 rpm for 3 min at room temperature.
7. Transfer the supernatant to a new 2.0 ml Eppendorf tube.
8. Add 350 µl of neutral phenol:chloroform:isoamyl alcohol (25:24:1; Sigma-Aldrich cat. no. P3803) and mix briefly for 10 sec.
9. Centrifuge at 13,000 rpm for 10 min at room temperature.
10. Transfer supernatant to a new 2.0 ml Eppendorf tube.
11. Add 300 µl of chloroform and mix briefly for 10 sec.
12. Centrifuge at 13,000 rpm for 10 min at room temperature.
13. Transfer supernatant to a new 1.5 ml Eppendorf tube.
14. Add 600 µl of pre-cooled isopropanol and mix. Keep on ice for 20 min at -20 °C.
15. Centrifuge at 13,000 rpm for 10 min at 4 °C.
16. Discard supernatant and wash with 300 µl 70% ethanol.
17. Centrifuge at 13,000 rpm for 10 min at room temperature.
18. Discard supernatant and let the sample dry for 15–30 min (do not dry too much) and dissolve in 50 µl Tris-EDTA buffer.
19. Keep at 4 °C overnight before use.

F.1.3 Isolation of *Streptomyces* genomic DNA

1. Start a culture of *S. leeuwenhoekii* C34 in 50 ml TSB medium at an appropriate temperature for the strain in use, 250 rpm, until reaching high cell density (usually about 48 h).
2. Aliquot the culture in 5 ml volumes in 15 ml conical centrifuge tubes and centrifuge at 3,500 rpm for 10 min.
3. Discard the supernatant and resuspend in 5 ml 20% glycerol and store at -20 °C until isolation of genomic DNA.
4. Thaw cells and centrifuge at 3,500 rpm for 10 min and discard glycerol.
5. Resuspend cell pellet in 5 ml SET buffer and add 150 µl of 50 mg/ml lysozyme and incubate at 37 °C for 2 h. Mix briefly every 30 min.
6. Add 300 µl of 20% SDS solution (from supplier).
7. Add 75 µl proteinase K (Fermentas E00491-1ML) and incubate at 60 °C for 1.5 h until the mixture becomes translucent.
8. Add 2 ml of 5 M NaCl and mix slowly. Here DNA will precipitate.
9. Fill up the conical centrifuge tube with chloroform (analytic grade).
10. Mix by inversion in a rotary mixer for 30 min.
11. Centrifuge the sample at 4,000 rpm for 20 min
12. Carefully collect the first phase of the solution with a snipped and sterile tip. Register the collected volume.

13. Add the same volume of isopropanol as that registered in the previous step and mix by inversion, briefly.
14. Add 1 ml of 70% ethanol to a sterile 2.0 ml Eppendorf tube.
15. With a Pasteur pipette make a hook to *fish out* the DNA. Transfer the DNA to the tube with ethanol.
16. Centrifuge at 4.000 rpm, 1 min.
17. Carefully pipet-out the supernatant. Dry the samples at 60 °C (do not dry too much).
18. Depending on the yield of the process, resuspend in 20−100 µl Tris-EDTA buffer.

F.2 Preparation of competent cells and transformation of *E. coli*

F.2.1 *E. coli* calcium chloride competent cells protocol

1. Start an overnight culture of *E. coli* TOP10 or DH5α in 5 ml of LB at 37 °C.
2. Inoculate 50 ml of fresh LB, in 250 Erlenmeyer flasks, with 1 ml (1/50) of *E. coli* from the overnight culture. Incubate at 37 °C and 250 rpm until an OD measured at 600 nm between 0.2-0.3 is reached.
3. Place culture on ice for 10 min.
4. Centrifuge culture in 50 ml conical centrifuge tubes at 3,000 rpm at 4 °C for 10 min.
5. Remove supernatant and resuspend cell pellet in 12.5 ml (1/4) 0.1 M $MgCl_2$ and place the resuspended pellet on ice for 5 min.
6. Centrifuge at 4 °C and 4,000 rpm for 10 min.
7. Remove supernatant and resuspend cell pellet in 2.5 ml (1/20) 0.1 M $CaCl_2$ and place the resuspended pellet on ice for 20 min.
8. Centrifuge at 4 °C and 4,000 rpm for 10 min. Remove supernatant and resuspend cell pellet in 1 ml (1/50) of 85% 0.1 M $CaCl_2$ solution plus 15% glycerol.

Note. The solution 85% 0.1 M $CaCl_2$ plus 15% glycerol contains, for culture of 50 ml of initial volume: 42.5 ml 0.1M $CaCl_2$ plus 7.5 ml glycerol 80%.

9. Aliquot the resulting suspension in 100 µl in sterile 0.6 ml Eppendorf tubes and store at -80 °C until use.

F.2.2 *E. coli* electrocompetent cells protocol

1. Start an overnight culture of *E. coli* strain in 5 ml of LB at 30 °C or 37 °C with (or without) relevant antibiotics and stir at 250 rpm.
2. Early the next morning, inoculate 50 ml of fresh LB, supplemented with (or without) relevant antibiotics, with 1 ml (1/50) of *E. coli* from the overnight culture. Incubate at 30 °C or 37 °C at 250 rpm until OD measured at 600 nm between 0.5−0.7 is reached. Place glycerol solution on ice for further use.
3. Place the culture on ice for 10 min.
4. Centrifuge culture in 50 ml conical centrifuge tubes at 4,000 rpm for 5 min at 4 °C.
5. Remove supernatant and resuspend cell pellet in 20 ml (1/2.5) 10% glycerol (kept on ice).
6. Centrifuge at 4 °C and 4,000 rpm for 5 min.

7. Remove supernatant and resuspend cell pellet in 20 ml (1/2.5) 10% glycerol (kept on ice).
8. Centrifuge at 4 °C and 4,000 rpm for 5 min.
9. Remove supernatant and resuspend cell pellet in 1 ml 10% glycerol (kept on ice).
10. Prepare 100 µl aliquots in 0.6 ml Eppendorf tubes (previously autoclaved) and store at -80 °C.

F.2.3 Chemical transformation of *E. coli*

1. Thaw calcium chloride competent *E. coli* TOP10 or DH5α stock on ice.
2. Mix 100 µl of calcium chloride competent *E. coli* with 10 µl of purified plasmid or ligation mixture in a 1.5 ml Eppendorf tube and keep on ice for 30 min. Use freeze tips.
3. Incubate in a water bath at 42 °C for 45 sec and then place on ice for 1 min.
4. Add 900 µl SOC medium and incubate at 37 °C at 250 rpm for 1 h.
5. Spread volumes of 200 µl, 150 µl and 90 µl on LA plates supplemented with a volume of carbenicillin, X-Gal and IPTG to reach a final concentration of 0.1 mg/ml, 0.1 mM and 40 µg/ml, respectively.

Note. Add IPTG, X-Gal and carbenicillin after autoclaving LA, when reaching about 40 °C.

6. Incubate plates at 37 °C overnight and screen for the desired transformant.

F.2.4 Transformation of *E. coli* by electroporation

1. Thaw electrocompetent *E. coli* TOP10 stock on ice.
2. Mix 100 µl of electrocompetent *E. coli* with 10 µl of purified plasmid or ligation mixture in a 1.5 ml Eppendorf tube and keep on ice for 30 min. Use freeze tips.
3. Place the mixture in a freeze electroporation cuvette (0.2 cm gap width) and carry out electroporation in BioRad GenePulser at 2,500 V. Use freeze cuvette.
4. Immediately add 900 µl ice-cold LB and incubate for 1 h at 37 °C.
5. Spread volumes of 200 µl, 150 µl and 90 µl on LA plates supplemented with relevant antibiotics.
6. Incubate at 37 °C overnight and screen for the desired transformant.

F.3 General procedures for cultivation and genetic manipulation of *Streptomyces* strains

F.3.1 Spore stock preparation

Protocol adapted from http://openwetware.org/wiki/Streptomyces:Protocols/Spore_Prep and Kieser et al., 2000.

1. Streak out a colony of *S. leeuwenhoekii* C34 or *S. coelicolor* onto fresh SFM agar plate, using a sterile cotton bud. Incubate at 30 °C until full sporulation is reached.

2. Prepare the filtration system. Stuff a 10 ml syringe with hydrophilic cotton as indicated in Figure F.1A. Introduce the plunger inside the barrel, leaving the plunger seal at the top of the barrel. Wrap the filtration system with aluminium foil and autoclave. Do not touch the adaptor throughout the process.
3. After the cells reached full sporulation, flood the plates with 2 ml 20% glycerol and gently rub the surface of the plate with a cell spreader made with a Pasteur pipette (Figure F.1B).
4. Transfer the solution (20% glycerol plus spores and debris) with a snipped tip into a clean 1.5 ml Eppendorf tube and apply vortex, briefly (Figure F.1C).

Optional. Place the Eppendorf tube in an ultrasonic bath sonicator at 40,000 Hz for 3 min.

5. Pour the contents of the Eppendorf tube into the filtration system (suggestion: place the syringe with its adaptor inside a clean 1.5 ml Eppendorf tube while the plunger is separated from the barrel).
6. Expel the solution from the filtration system into a sterile Eppendorf tube.
7. Store the spore stock at -20 °C.

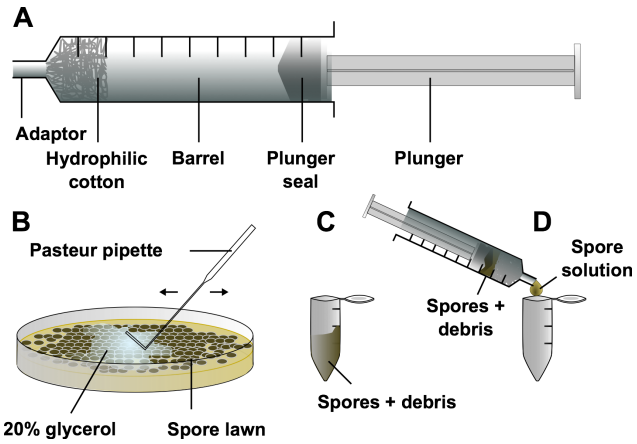


Figure F.1: Schematic representation of the process for preparing spore stocks of *S. leeuwenhoekii* C34 or *S. coelicolor*. (**A**) Filtration system; (**B**) separation of the spores from the surface of the plate; (**C**) temporary storage of spores plus debris solution; (**D**) filtration of spore solution.

F.3.2 Preparation of seed cultures of *Streptomyces* strains

1. Before starting, coat the internal wall of a 250 ml Erlenmeyer flask with Sylon CTTM (5% dimethyldichlorosilane in Toluene; Supelco catalogue number 33065-U) and allow to dry overnight. This step is required to deter the mycelium from sticking to the glass.
2. Place a stainless steel spring at the bottom of each flask to improve aeration. Autoclave the flask.
3. Add 25 ml of liquid medium to the flasks and inoculate with $1{\cdot}10^8$ spores. Modified ISP2 was used as chaxamycin production medium for *S. leeuwenhoekii* C34 (Rateb et al., 2011a) and R3 medium was used for *S. coelicolor*.
4. Incubate at 30 °C for 24 h or 48 h at 250 rpm.
5. Pour the contents of the flask into a 50 ml conical centrifuge tube and centrifuge at 4,000 rpm at room temperature. Discard the supernatant.
6. Wash the cell pellet twice in normal saline solution and resuspend in fresh production medium to begin the experiment.

F.3.3 Conjugation between *E. coli* ET12567 strains and *S. leeuwenhoekii* C34

1. Set up an overnight culture of *E. coli* ET12567, carrying either pUZ8002 or pR9406 (helper plasmids) and the plasmid to be transferred, in 5 ml LB supplemented with relevant antibiotics. Leave the culture at 37 °C and 250 rpm.
2. Early the next morning, sub-culture *E. coli* in 10 ml LB supplemented with relevant antibiotics at 37 °C and 250 rpm until reaching an OD measured at 600 nm of 0.3−0.4.
3. Centrifuge the culture at room temperature and 3,000 rpm, discard the supernatant and wash the cell pellet in 10 ml LB. Repeat this process twice. Resuspend the cell pellet in 0.3 ml LB.
4. Heat-shock $1{\cdot}10^8$ spores of *S. leeuwenhoekii* C34 in 0.5 ml of 2xYT medium (recipe in Appendix E.2) at 50 °C for 10 min. Cool to room temperature.
5. Centrifuge at room temperature and 10,000 rpm for 5 min and discard supernatant.
6. Resuspend the spores with the *E. coli* suspension and mix gently.
7. Spread 0.1 ml of the conjugation solution with a glass bacterial cell spreader on SFM agar medium, supplemented with 10 mM $CaCl_2$ and 10 mM $MgCl_2$ and without antibiotics. Let the surface of the plates dry at room temperature.
8. Incubate for 20−24 h at 30 °C.
9. Overlay with 1 ml of nalidixic acid plus relevant antibiotic solution, per plate (see Table F.1; Kieser et al., 2000). For example when overlaying with apramycin, the solution would be: 25 µl apramycin plus 20 µl nalidixic acid plus 955 µl sterile water.
10. Let the surface of the plate to dry and incubate at 30 °C.

Controls. Spread 0.1 ml of the conjugation solution onto SFM agar medium and overlay the next day with just nalidixic acid.

F.3.4 Transference of PAC DNA from *E. coli* ET12567 into *S. coelicolor* (Tri-parental mating)

Step 1 Tri-parental mating using three *E. coli* strains to mobilise PAC clones from *E. coli* DH10B into *E. coli* ET12567.

1. Sub-culture overnight cultures of *E. coli* DH10B containing the PAC clone (pESAC13 derivative, kanamycin resistance in *E. coli*, thiostrepton resistance in *Streptomyces*), *E. coli* TOP10 containing the self-transmissible helper plasmid pR9604 (carbenicillin resistant) and *E. coli* ET12567 (recipient) (chloramphenicol resistant) into fresh 10 ml LB and grow them into exponential phase.

Note The strains have different growth rates, so carry out a trial experiment under local conditions first.

2. Wash the cultures twice with fresh LB liquid medium (using centrifugation to pellet the cells), to remove residual antibiotics.
3. Resuspend each pellet in 50 µl of fresh LB.
4. Pour LA plates lacking antibiotics for the tri-parental mating. Pilli are very fragile and it has been mentioned that they shear off in liquid culture.
5. Spot 10 to 20 µl of each strain onto the same location on the LA plate so that the three strains are mixed together. Wait for the spots to dry in and incubate at 37 °C overnight for tri-parental conjugation.
6. Next day, prepare LA plates containing chloramphenicol, kanamycin and carbenicillin to select for *E. coli* ET12567 derivatives containing the PAC clone and the helper plasmid (pR9604).

7. Streak the overnight incubated spots from the LA conjugation plates onto the LA plates containing kanamycin, carbenicillin and chloramphenicol to obtain single colony exconjugants of *E. coli* ET12567 containing the pESAC13 derivative and pR9604.
8. After overnight incubation, inoculate a few of the resulting colonies into LB containing chloramphenicol, kanamycin and carbenicillin and carry out PCR analysis to confirm the presence of the PAC clone.

Step 2 Conjugation between *E. coli* ET12567 containing pESAC13 derivatives and pR9604, and *S. coelicolor* strains.

9. Follow the conjugation protocol described in Appendix F.3.3, but instead of plating out the conjugation mixture on a SFM agar plate, use R2-S agar medium.
10. After flooding the R2-S agar conjugation plate with thiostrepton and nalidixic acid, wait for a few days to observe putative exconjugants. Streak the colonies onto DNA agar supplemented with thiostrepton plus nalidixic acid. If they survive, streak out again onto SFM agar supplemented with nalidixic acid.
11. Carry out PCR analysis to confirm that the entire PAC clone has been transferred to the *S. coelicolor* recipient, which is resistant to thiostrepton.

Table F.1: Concentration of antibiotics used to select for transformed *S. leeuwenhoekii* C34 and *S. coelicolor* strains. Preparation of antibiotic solutions are listed in Appendix E.3.1.

Antibiotic (stock concentration mg/ml)	Final concentration of antibiotic in 25 ml agar plates (µg/ml) after overlaying	Amount of stock solution (µl)
Apramycin (50)	50	25
Hygromycin B (50)	40	20
Kanamycin (50)	50	25[1]
Thiostrepton (50)	60	30[2]
Nalidixic acid (25)	20	20

[1] In this work, this concentration was only used for *S. leeuwenhoekii* C34.
[2] In this work, this concentration was only used for *S. coelicolor*.

F.4 Bioassays against *Bacillus subtilis* and *Micrococcus luteus*

1. Set up a overnight culture of *B. subtilis* and *M. luteus* in 10 ml LB. Incubate at 37 °C at 250 rpm.
2. Very early the next morning, take 1 ml of the previous culture and inoculate 50 ml LB. Do this for each microorganism. Incubate at 30 °C at 250 rpm until an OD measured at 600 nm of 0.6 is reached. It takes about 4 h for *B. subtilis* and about 7 h for *M. luteus*, although it depends on the density of the pre-inoculum.
3. Once the desired OD measured at 600 nm is reached, centrifuge the cultures and resuspend in 3 ml LB. Measure OD measured at 600 nm of the concentrated cells.
4. Melt LA (previously autoclaved) and let it cool to about 40 °C; then, add a volume of either *B. subtilis* or *M. luteus* such that the OD measured at 600 nm is 0.0125 in the medium. Quickly, mix by gentle manual stirring since agar would start to solidify.
5. Pour LA with *B. subtilis* or *M. luteus* onto square Petri dishes and wait until the agar becomes solid.
6. Using sterile forceps, take a tip and with the back of it make wells on the agar; these wells will contain the sample for testing antimicrobial activity.

7. Fill each well with 50 µl of sample. Use either, 100 µg/ml apramycin or 100 µg/ml carbenicillin as positive control; use culture medium as negative control.
8. Refrigerate the plates at 4 °C for 45 min−1 h to allow liquid sample to diffuse through the agar. The surface of the plate must always be facing upwards.
9. Incubate the plates for 1 or 2 days at 30 °C, plate must be facing upwards.

F.5 Measurement of chaxamycins in liquid cultures of *S. leeuwenhoekii* C34 and *S. coelicolor* M1650

Chaxamycin A, chaxamycin B and chaxamycin C standards were kindly provided by Dr. Mostafa Rateb and Prof. Marcel Jaspars (University of Aberdeen, Scotland).

1. Samples of 1 ml of liquid culture of *S. leeuwenhoekii* C34 or *S. coelicolor* strains were withdrawn from the flasks at different times and collected in 1.5 ml Eppendorf tube.
2. Samples were centrifuged twice at 13,000 rpm at 4 °C for 10 min, and the supernatant was carefully transferred into new Eppendorf tubes. An aliquot of 0.5 ml of the supernatant was placed in an HPLC vial.

Optional. The mycelial pellet from the first centrifugation step was extracted with 0.5 ml of methanol, centrifuged at 13,000 rpm at 4 °C for 10 min, and 0.3 ml of the methanol extract was mixed with 0.3 ml water in an HPLC vial.

3. Samples were analysed on a Shimadzu LC-MS system equipped with a NexeraX2 liquid chromatograph (LC30AD) fitted with a Prominence photo diode array detector (SPD-M20A) and an LCMS-IT-ToF mass spectrometer.
4. Five µl of sample was injected in a Kinetex XB-C18 2.6 µm 100 Å 50x2.10 mm column (part no. 00B-4496-AN, Phenomenex, USA) fitted with a KrudKatcher Ultra HPLC in-line filter (part no. AF0-8497, Phenomenex, USA).
5. The elution flow rate was set to 0.6 ml per min with a gradient of 0.1% v/v formic acid in water (mobile phase A) and methanol (mobile phase B) as follows (minute, % of B reached at that minute): 0 min, 2% B; 1 min, 2% B; 8 min, 100% B; 9.3 min, 100% B; 9.5 min, 2% B; 11.2 min, 2% B for equilibration. The column was kept at 40 °C.
6. Chaxamycin were detected through mass spectrometry in negative ionization mode (more sensitivity for chaxamycin detection was observed in this mode than in positive ionization mode).
7. Purified chaxamycin A, chaxamycin B and chaxamycin C standards were used in each run.
8. Data acquisition and analysis were performed with LCMSsolutions version 3 (Shimadzu); spectra visualisation was also performed with Mass++ version 2.7.2 (http://www.first-ms3d.jp/english/). The $[M - H]^-$ ions detected for the chaxamycin species were: chaxamycin A *m/z* 638.29 $[M - H]^-$; chaxamycin B *m/z* 622.29 $[M - H]^-$; chaxamycin C *m/z* 654.29 $[M - H]^-$; chaxamycin D *m/z* 682.29 $[M - H]^-$.

Appendix G

Supplementary information of Chapter One: Establishment of practical procedures for growth and genetic manipulation of *Streptomyces leeuwenhoekii* C34

Table G.1: Putative DNA sequences of *attB* sites in *S. leeuwenhoekii* C34.

attB site	DNA sequence 5'→3'
ΦC31-Sco	CGGTGCGGGTGCCAGGGCGTGCCC**TT**GGGCTCCCCGGGCGCGTACTCCAC
ΦC31-Sle	CGGTGGGGGTGCCAGGGGGTGCCC**TT**GGGCTCTCCGGGCGCGTACTCCAC
ΦBT1-Sco	TCCTTGACCAGGTTTTTGACGAAA**GT**GATCCAGATGATCCAGCTCCACAC
ΦBT1-Sle	TCCTTGATCAGATTTTTGACGAAA**GT**GATCCAGATGACCCAGCTCCACAC

attB recognised by ΦC31 (Rausch and Lehmann, 1991) and ΦBT1 (Gregory et al., 2003) integrases in *S. coelicolor* (Sco) and putatively in *S. leeuwenhoekii* C34 (Sle). Differences are underlined.

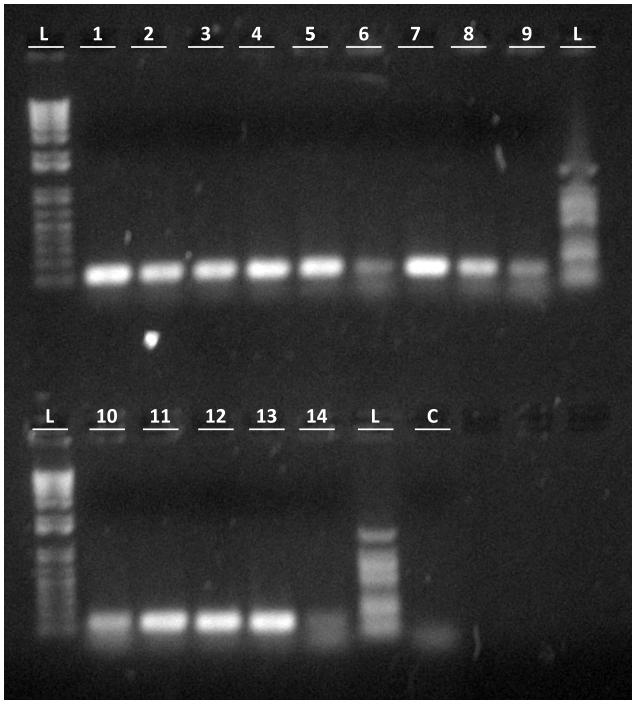


Figure G.1: Verification of insertion of pSET152 in the genome of *S. leeuwenhoekii* strain M1660, obtained from optimised conjugation conditions. Lane L is 100 bp DNA ladder (cat. no. N3231S); lanes 1−14 are *S. leeuwenhoekii* M1660 exconjugants; lane C is genomic DNA from *S. leeuwenhoekii* C34 wild-type.

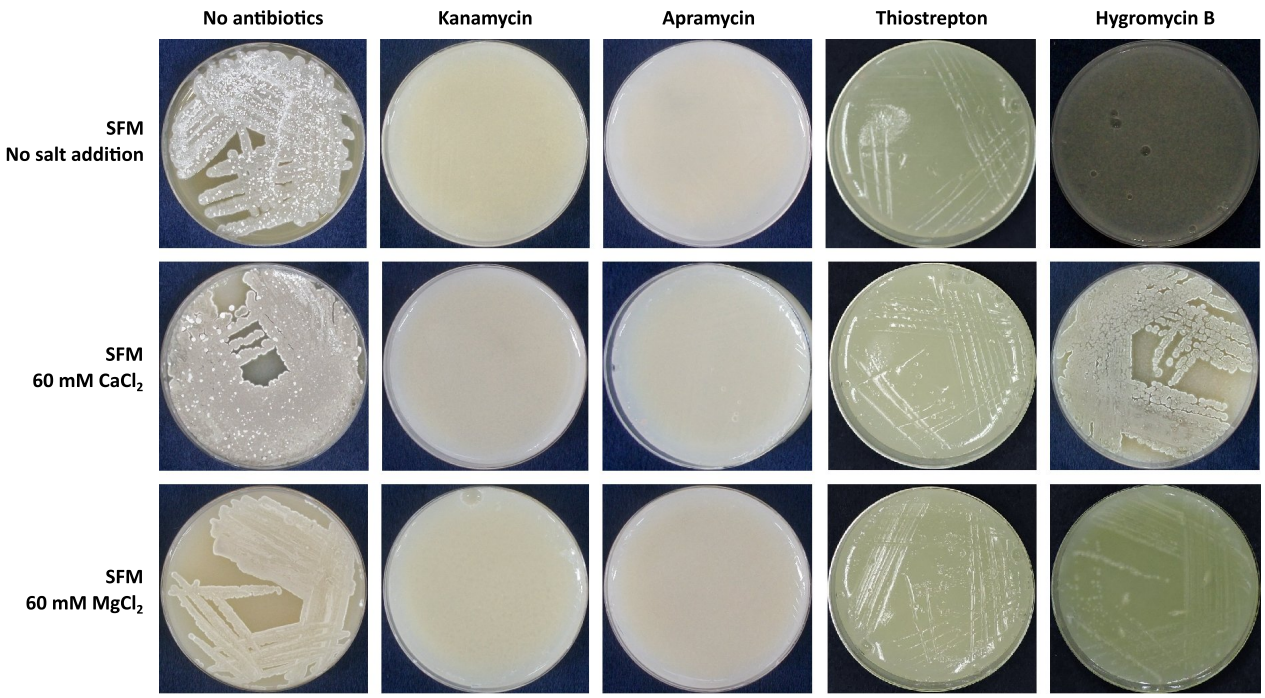


Figure G.2: Evaluation of the effectiveness of antibiotics to prevent growth of *S. leeuwenhoekii* C34 in presence of either 60 mM $CaCl_2$ or 60 mM $MgCl_2$. *S. leeuwenhoekii* C34 was cultivated on SFM supplemented with 60 mM $CaCl_2$ plus antibiotic at a concentration typically used: 100 µg/ml kanamycin, 50 µg/ml apramycin, 50 µg/ml thiostrepton and 50 µg/ml hygromycin B; plates were incubated at 30 °C for three days.

Table G.2: rRNA methyltransferases of *S. leeuwenhoekii* C34.

Protein code	Organism	Protein name (NCBI accession)	Domains (Pfam accession)
Tsr	*S. azureus*	23S rRNA [adenosine(1067)-2'-*O*]-methyltransferase (P18644)	TSNR_N (PF04705) SpoU_methylase (PF00588)
Sle_55160	*S. leeuwenhoekii* C34	Uncharacterized tRNA/rRNA methyltransferase (CQR64973)	SpoU_sub_bind (PF08032) SpoU_methylase (PF00588)
Sle_38790	*S. leeuwenhoekii* C34	Uncharacterized tRNA/rRNA methyltransferase (CQR63338)	SpoU_sub_bind (PF08032) SpoU_methylase (PF00588)
Sle_55570	*S. leeuwenhoekii* C34	Uncharacterized tRNA/rRNA methyltransferase (CQR65014)	SpoU_methylase (PF00588)
Sle_35420	*S. leeuwenhoekii* C34	rRNA methylase (CQR63003)	SpoU_methylase (PF00588)

Table G.3: APHs of *S. leeuwenhoekii* C34. Pfam accession of the APH domain: PF01636.

Protein code	Organism	Protein name (NCBI accession)	Domains (CDD accession)
Sle_01670	*S. leeuwenhoekii* C34	APH homologue to Hyg21 of the hygromycin A biosynthesis gene cluster of *S. hygroscopicus* (CQR59629) (Dhote et al., 2008)	AAA_33 (PF13671)
Sle_14060	*S. leeuwenhoekii* C34	Aminoglycoside phosphotransferase (CQR60868)	PKc_like super family (COG3173)
Sle_15890	*S. leeuwenhoekii* C34	Aminoglycoside phosphotransferase (CQR61051)	PKc_like super family (COG3173)
Sle_24140	*S. leeuwenhoekii* C34	Aminoglycoside phosphotransferase (CQR61875)	APH_ChoK_like_1 (CD05155) PKc_like super family (COG3173)
Sle_54290	*S. leeuwenhoekii* C34	Aminoglycoside phosphotransferase (CQR64886)	PKc_like super family (COG3173) ACAD10_N-like (CD05154)

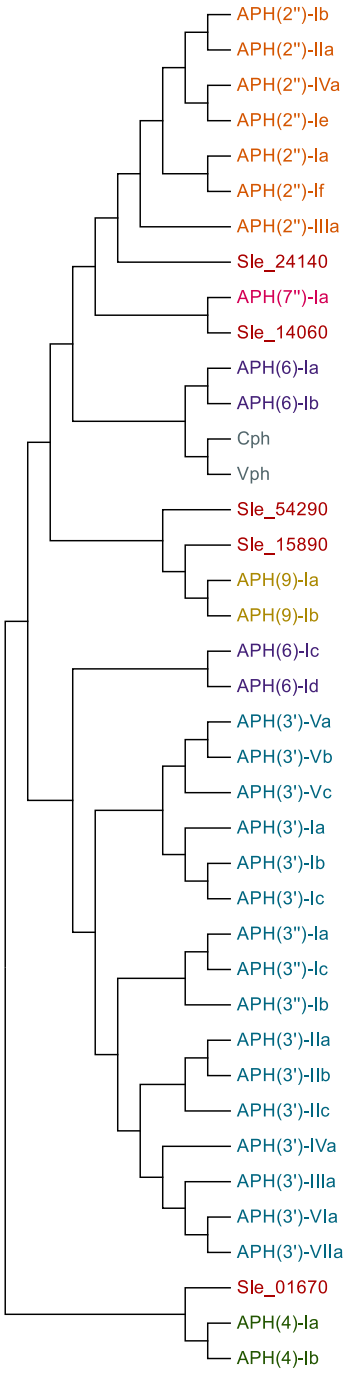


Figure G.3: Phylogenetic tree of protein sequences of characterised APHs and those homologues encoded in the genome of *S. leeuwenhoekii* C34. Sequences of characterised APHs were obtained from Wright and Thompson, 1999 and Ramirez and Tolmasky, 2010. Neighbour-joining tree was created with MEGA6 (Tamura et al., 2013), using maximum likelihood as statistical test (other parameters: Gap/missing data treat = partial deletion; site coverage cutoff (%) = 95; bootstrap method with 1000 replications; rates among sites = Gamma distributed with Gamma-parameter of 1; Hall, 2013). Image of the phylogenetic tree was generated with Interactive Tree Of Life (iTOL) on-line tool (Letunic and Bork, 2011) and edited in Inkscape. Sle_14060, Sle_15890, Sle_24140 and Sle_54290 belong to the APH family enzymes of aminoglycoside phosphotransferase (Pfam accession: PF01636) and contain the PKc_like super family domain that is a protein-kinase catalytic domain (CDD accession: CD05120). The list of sequences used can be found on the *next page*.

List of APH enzymes used to build a phylogenetic tree (Figure Appendix G.3).

AHP	Organism	Protein name	Protein accession
APH(4)-Ia	*Escherichia coli*	Hygromycin-B 4-*O*-kinase	P00557
APH(4)-Ib	*Burkholderia pseudomallei*	Hygromycin phosphotransferase	Q52501
APH(6)-Ia	*Streptomyces griseus*	Streptomycin 6-kinase	P08077
APH(6)-Ib	*Streptomyces glaucescens*	Streptomycin 6-kinase	P18622
APH(6)-Ic	*Klebsiella pneumoniae*	Streptomycin 3”-kinase	P13082
APH(6)-Id	*Escherichia coli*	Aminoglycoside 6”-phosphotransferase	Q7ATJ6
APH(9)-Ia	*Legionella pneumophila*	Spectinomycin phosphotransferase	O06916
APH(9)-Ib	*Streptomyces netropsis*	Spectinomycin phosphotransferase	P72441
APH(3’)-Ia	*Escherichia coli*	Aminoglycoside 3’-phosphotransferase	P00551
APH(3’)-Ib	*Escherichia coli*	Aminoglycoside 3’phosphotransferase	P14509
APH(3’)-Ic	*Klebsiella pneumoniae*	Aminoglycoside 3’-phosphotransferase	Q79AN4
APH(3’)-IIa	*Klebsiella pneumoniae*	Aminoglycoside 3’-phosphotransferase	P00552
APH(3’)-IIb	*Pseudomonas aeruginosa*	Aminoglycoside 3’-phosphotransferase	Q9HWR2
APH(3’)-IIc	*Stenotrophomonas maltophilia*	Aminoglycoside 3’-phosphotransferase	E5LCR3
APH(3’)-IIIa	*Enterococcus faecalis*	Aminoglycoside 3’-phosphotransferase	P0A3Y5
APH(3’)-IVa	*Bacillus circulans*	Aminoglycoside 3’-phosphotransferase	P00553
APH(3’)-Va	*Streptomyces fradiae*	Aminoglycoside 3’-phosphotransferase	P00555
APH(3’)-Vb	*Streptomyces ribosidificus*	Aminoglycoside 3’-phosphotransferase	P13250
APH(3’)-Vc	*Micromonospora chalcea*	Aminoglycoside-*O*-phosphoryltransferase	Q53558
APH(3’)-VIa	*Acinetobacter baumannii*	Aminoglycoside 3’-phosphotransferase	P09885
APH(3’)-VIb	*Klebsiela pneumoniae*	-	N.a.[1]
APH(3’)-VIIa	*Campylobacter jejuni*	Aminoglycoside 3’-phosphotransferase	P14508
APH(2”)-Ia	*Staphylococcus aureus*	Bifunctional AAC/APH	P0A0C0
APH(2”)-Ib	*Enterococcus faecium*	Aminoglycoside phosphotransferase	Q9EVD7
APH(2”)-IIa	*Escherichia coli*	Aminoglycoside phosphotransferase	Q93ET9
APH(2”)-IIIa	*Enterococcus gallinarum*	Gentamicin resistance protein	P96762

Continued. List of APH enzymes used to build a phylogenetic tree.

AHP	Organism	Protein name	Protein accession
APH(2")-IVa	*Enterococcus casseliflavus*	APH(2")-Id	O68183
APH(2")-Ie	*Enterococcus casseliflavus*	Aph(2")-Ie	Q58M16
APH(2")-If	*Campylobacter jejuni*	Bifunctional aminoglycoside modifying enzyme	Q4VR94
APH(3")-Ia	*Streptomyces griseus*	Streptomycin 3"-kinase	P18150
APH(3")-Ib	*Pseudomonas syringae*	APH(3") streptomycin phosphotransferase	Q7BKK6
APH(3")-Ic	*Mycobacterium fortuitum*	APH(3")-Ic	Q0ZNJ3
APH(7")-Ia	*Streptomyces hygroscopicus*	Hygromycin-B 7"-*O*-kinase	P09979
Vph[2]	*Streptomyces vinaceus*	Viomycin phosphotransferase	P18623
Cph[2]	*Streptomyces capreolus*	Capreomycin phosphotransferase	Q53826
Sle_01670[3]	*Streptomyces leeuwenhoekii*	Putative hygromycin A phosphotransferase	A0A0F7VKS8
Sle_14060[4]	*Streptomyces leeuwenhoekii*	Putative aminoglycoside phosphotransferase	A0A0F7VQN4
Sle_15890[4]	*Streptomyces leeuwenhoekii*	Putative aminoglycoside phosphotransferase	A0A0F7VVP4
Sle_24140[4]	*Streptomyces leeuwenhoekii*	Putative aminoglycoside phosphotransferase	A0A0F7VPY5
Sle_54290[4]	*Streptomyces leeuwenhoekii*	Putative aminoglycoside phosphotransferase	A0A0F7W6M0

[1] Sequence not available.
[2] Other type of APH.
[3] Putative hygromycin A APH, based on homology with Hyg21, an *O*-phosphotransferase present in *S. hygroscopicus* NRRL 2388 (Dhote et al., 2008).
[4] Putative APH encoded in the genome of *S. leeuwenhoekii* C34.

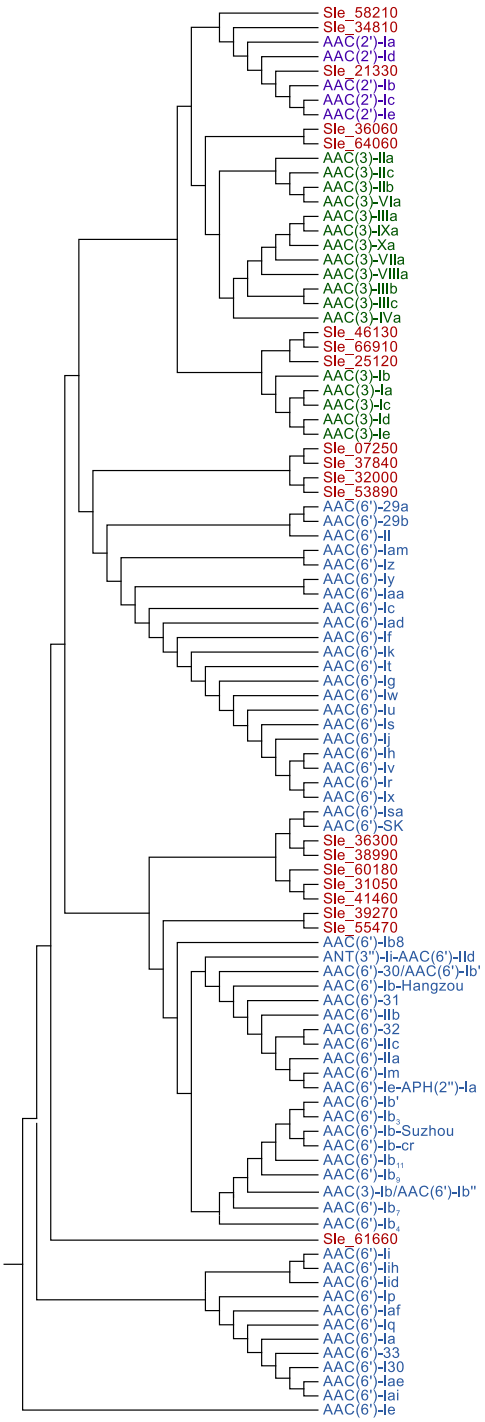


Figure G.4: Phylogenetic tree of protein sequences of characterised AAC and those homologues encoded in the genome of *S. leeuwenhoekii* C34. Sequences of characterised AACs were obtained from Ramirez and Tolmasky, 2010. The sequences of proteins of *S. leeuwenhoekii* C34 were selected based on the presence of the following Pfam domains (accession), found in AACs: acetyltransf_1 (PF00583), acetyltransf_3 (PF13302), acetyltransf_7 (PF13508), acetyltransf_8 (PF13523), acetyltransf_10 (PF13673) and domain of unknown function 4111 (PF13427). Neighbour-joining tree was created with MEGA6 (Tamura et al., 2013), using maximum likelihood as statistical test (other parameters: Gap/missing data treat = partial deletion; site coverage cutoff (%) = 95; bootstrap method with 1000 replications; rates among sites = Gamma distributed with Gamma-parameter of 1; Hall, 2013). Image of the phylogenetic tree was generated with Interactive Tree Of Life (iTOL) on-line tool (Letunic and Bork, 2011) and edited in Inkscape. The list of sequences used can be found on the *next page*.

List of AAC enzymes used to build a phylogenetic tree (Figure Appendix G.4).

AAC	Organism	Protein name	Protein accession
AAC(3)-Ia	*Escherichia coli*	*N*-acetyltransferase	Q849Y5
AAC(3)-Ib	*Pseudomonas aeruginosa*	Aminoglycoside 3'-*N*-acetyltransferase	Q51407
AAC(3)-Ic	*Pseudomonas aeruginosa*	*N*-acetyltransferase	Q83V16
AAC(3)-Id	*Salmonella newport*	Aminoglycoside acetyltransferase	Q6SIX1
AAC(3)-Ie	*Salmonella enterica*	Aminoglycoside (3) acetyltransferase	Q6S741
AAC(3)-IIa	*Escherichia coli*	Aminoglycoside-(3)-*N*-acetyltransferase	A1YKW2
AAC(3)-IIb	*Serratia marcescens*	Aminoglycoside *N*-(3')-acetyltransferase III	Q01515
AAC(3)-IIc	*Escherichia coli*	Aminoglycoside *N*-(3')-acetyltransferase	Q03634
AAC(3)-IIIa	*Pseudomonas aeruginosa*	Aminoglycoside *N*-(3')-acetyltransferase III	P29808
AAC(3)-IIIb	*Pseudomonas aeruginosa*	Aminoglycoside *N*-(3')-acetyltransferase III	Q51405
AAC(3)-IIIc	*Pseudomonas aeruginosa*	Aminoglycoside *N*-(3')-acetyltransferase III	Q51406
AAC(3)-IVa	*Salmonella* sp.	Aminoglycoside *N*-(3')-acetyltransferase IV	P08988
AAC(3)-VIa	*Enterobacter cloacae*	Aminoglycoside *N*-(3')-acetyltransferase	Q47030
AAC(3)-VIIa	*Streptomyces rimosus* subsp. *paromomycinus*	Aminoglycoside *N*-(3')-acetyltransferase VII	P30180
AAC(3)-VIIIa	*Streptomyces fradiae*	Aminoglycoside *N*-(3')-acetyltransferase VIII	P29809
AAC(3)-IXa	*Micromonospora chalcea*	Aminoglycoside *N*-(3')-acetyltransferase IX	P29810
AAC(3)-X	*Streptomyces griseus*	Aminoglycoside 3-*N*-acetyltransferase	Q54216
AAC(2')-Ia	*Providencia stuartii*	Aminoglycoside 2'-*N*-acetyltransferase	Q52424
AAC(2')-Ib	*Mycobacterium fortuitum*	Aminoglycoside 2'-*N*-acetyltransferase	Q49157
AAC(2')-Ic	*Mycobacterium tuberculosis*	Aminoglycoside 2'-*N*-acetyltransferase	P9WQG8
AAC(2')-Id	*Mycobacterium smegmatis*	Aminoglycoside 2'-*N*-acetyltransferase	P94968
AAC(2')-Ie	*Mycobacterium leprae*	Aminoglycoside 2'-*N*-acetyltransferase	Q9CD24
AAC(6')-Iam	*Stenotrophomonas maltophilia*	Aminoglycoside *N*-(6')-acetyltransferase I	B2FQI3
AAC(6')-Ia	*Citrobacter koseri*	Aminoglycoside *N*-(6')-acetyltransferase I	P10051
AAC(6')-Ib	*Klebsiella pneumoniae*	Aminoglycoside *N*-(6')-acetyltransferase I	P19650
AAC(6')-Ib'	*Pseudomonas fluorescens*	Aminoglycoside *N*-(6')-acetyltransferase	Q79E71
AAC(6')-Ic	*Serratia marcescens*	Aminoglycoside *N*-(6')-acetyltransferase I	Q54441
AAC(6')-Ie	*Enterococcus faecalis*	Aac(6')-Ie-aph(2")-Ia protein	A0A0C6DW77
AAC(6')-If	*Enterobacter cloacae*	Aminoglycoside *N*-(6')-acetyltransferase I	Q03635
AAC(6')-Ig	*Acinetobacter haemolyticus*	Aminoglycoside *N*-(6')-acetyltransferase I	Q44057
AAC(6')-Ih	*Acinetobacter baumannii*	Aminoglycoside *N*-(6')-acetyltransferase I	Q43899
AAC(6')-Ii	*Enterococcus faecium*	*N*-acetyltransferase	Q47764
AAC(6')-Ij	*Acinetobacter* sp. genomospecies 13	Aminoglycoside *N*-(6')-acetyltransferase I	Q44245
AAC(6')-Ik	*Acinetobacter* sp. CIP-A165	Aminoglycoside *N*-(6')-acetyltransferase I	Q44246

Continued. List of AAC enzymes used to build a phylogenetic tree.

AAC	Organism	Protein name	Protein accession
AAC(6')-Ip	*Citrobacter freundii*	Aminoglycoside acetyltransferase (6') I	P94629
AAC(6')-Iq	*Klebsiella pneumoniae*	Aminoglycoside 6'-*N*-acetyltransferase	O68484
AAC(6')-Im	*Escherichia coli*	Aminoglycoside acetyltransferase	Q93ET8
AAC(6')-Il	*Enterobacter aerogenes*	Aminoglycoside *N*-(6')-acetyltransferase I	P50858
AAC(6')-Ir	*Acinetobacter* sp. genomospecies 14	Aminoglycoside *N*-(6')-acetyltransferase I	O52241
AAC(6')-Is	*Acinetobacter variabilis*	Aminoglycoside *N*-(6')-acetyltransferase I	O52242
AAC(6')−Isa	*Streptomyces albulus*	Kanamycin acetyltransferase	Q764C8
AAC(6')-It	*Acinetobacter* sp. genomospecies 16BJ	Aminoglycoside *N*-(6')-acetyltransferase I	O52243
AAC(6')-Iu	*Acinetobacter* sp. genomospecies 17BJ	Aminoglycoside *N*-(6')-acetyltransferase I	P00557
AAC(6')-Iv	*Acinetobacter* sp. 631	Aminoglycoside *N*-(6')-acetyltransferase I	O52245
AAC(6')-Iw	*Acinetobacter* sp. 640	Aminoglycoside *N*-(6')-acetyltransferase I	O52246
AAC(6')-Ix	*Acinetobacter* sp. BM2722	Aminoglycoside *N*-(6')-acetyltransferase I	O52247
AAC(6')-Iy	*Salmonella enteritidis*	Aminoglycoside *N*-(6')-acetyltransferase I	Q9R381
AAC(6')-Iz	*Stenotrophomonas maltophilia*	Aminoglycoside *N*-(6')-acetyltransferase I	Q9RBW7
AAC(6')-Iad	*Salmonella typhimurium*	Aminoglycoside *N*-(6')-acetyltransferase I	Q75ZP9
AAC(6')-Iae	*Pseudomonas aeruginosa*	Aminoglycosides acetyltransferase	Q76FR4
AAC(6')-Iaf	*Pseudomonas aeruginosa*	Aminoglycoside 6'-*N*-acetyltransferase	C4TGI5
AAC(6')-Iai	*Pseudomonas aeruginosa*	*N*-acetyltransferase	B6D9B6
AAC(6')-Ib_3	*Pseudomonas aeruginosa*	6'-*N*-acetyltransferase	Q7BH78
AAC(6')-Ib_4	*Serratia* sp.	Aminoglycoside 6'-*N*-acetyltransferase	Q79DM5
AAC(6')-Ib_7	*Enterobacter cloacae*	Aminoglycoside 6'-*N*-acetyltransferase	Y11946[1]
AAC(6')-Ib_8	*Enterobacter cloacae*	Aminoglycoside 6'-*N*-acetyltransferase	Y11947[1]
AAC(6')-Ib_9	*Pseudomonas aeruginosa*	Aminoglycoside 6'-*N*-acetyltransferase	Q9R977
AAC(6')-Ib_{11}	*Salmonella typhimurium*	Aminoglycoside 6'-*N*-acetyltransferase	Q8GLI5
AAC(6')-29a	*Pseudomonas aeruginosa*	6'-*N*-aminoglycoside acetyltransferase I	Q9AIK9
AAC(6')-29b	*Pseudomonas aeruginosa*	6'-*N*-aminoglycoside acetyltransferase	Q9AIK8
AAC(6')-31	*Pseudomonas putida*	6'-*N*-aminoglycoside acetyltransferase	Q0VTT5
AAC(6')-32	*Pseudomonas aeruginosa*	6'-*N*-aminoglycoside acetyltransferase	A6N6U3
AAC(6')-33	*Pseudomonas aeruginosa*	Aminoglycoside 6'-*N*-acetyltransferase	C7F6Y4
AAC(6')-I30	*Salmonella enterica*	6'-*N*-aminoglycoside acetyltransferase	Q7WTV9
AAC(6')-Iid	*Enterococcus durans*	Aminoglycoside acetyltransferase	Q70E71
AAC(6')-Iih	*Enterococcus hirae*	Aminoglycoside acetyltransferase	Q70E72
AAC(6')-Ib-Suzhou	*Enterobacter cloacae*	Aminoglycoside acetyltransferase	A7UMS1

Continued. List of AAC enzymes used to build a phylogenetic tree.

AAC	Organism	Protein name	Protein accession
AAC(6')-Ib-Hangzou	*Acinetobacter baumannii*	Aminoglycoside acetyltransferase	B8Y9I3
AAC(6')−SK	*Streptomyces kanamyceticus*	Aminoglycoside 6'-*N*-acetyltransferase	Q75PS3
AAC(6')-IIa	*Pseudomonas aeruginosa*	6'-*N*-acetyltransferase	Q51409
AAC(6')-IIb	*Pseudomonas fluorescens*	6'-*N*-acetyltransferase	Q51734
AAC(6')-IIc	*Enterobacter cloacae*	6'-*N*-acetyltransferase	Q6DT23
AAC(6')-Ib-cr	*Escherichia coli*	6'-*N*-acetyltransferase	Q6SJ71
AAC(6')-Ie-APH(2")-Ia	*Staphylococcus aureus*	Bifunctional AAC/APH	P0A0C1
ANT(3")-Ii-AAC(6')-IId	*Serratia marcescens*	Aminoglycoside (3") adenylyltransferase-aminoglycoside (6') acetyltransferase	Q8VQN7
AAC(6')-30/AAC(6')-Ib'	*Pseudomonas aeruginosa*	AAC(6')-30/AAC(6')-Ib'	Q70E84
AAC(3)-Ib/AAC(6')-Ib"	*Pseudomonas aeruginosa*	Aminoglycoside 3-N-acetyltransferase/aminoglycoside 6'-*N*-acetyltransferase	Q8RPU6
Sle_07250[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7VP36
Sle_21330[2]	*Streptomyces leeuwenhoekii*	Aminoglycoside 2'-*N*-acetyltransferase	A0A0F7VZH8
Sle_25120[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7VU34
Sle_31050[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7W0Q9
Sle_32000[2]	*Streptomyces leeuwenhoekii*	Uncharacterized *N*-acetyltransferase YjdG	A0A0F7W0Z7
Sle_34810[2]	*Streptomyces leeuwenhoekii*	GCN5-Related *N*-Acetyltransferase	A0A0F7VZH8
Sle_36060[2]	*Streptomyces leeuwenhoekii*	GCN5-Related *N*-Acetyltransferase	A0A0F7W234
Sle_36300[2]	*Streptomyces leeuwenhoekii*	Protein PhnO	A0A0F7VTR9
Sle_37840[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7VXQ1
Sle_38990[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7VXZ7
Sle_39270[2]	*Streptomyces leeuwenhoekii*	Uncharacterized *N*-acetyltransferase YjdG	A0A0F7W2Z8
Sle_41460[2]	*Streptomyces leeuwenhoekii*	Blasticidin-*S* acetyltransferase	A0A0F7VVA0
Sle_46130[2]	*Streptomyces leeuwenhoekii*	UPF0256 protein SAV_5428	A0A0F7W4X6
Sle_53890[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7W6J7
Sle_55470[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7VWR6
Sle_58210[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase ECA0875	A0A0F7W5Y4
Sle_60180[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase ECA0875	A0A0F7VY95
Sle_61660[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase ECA0875	A0A0F7W8N0
Sle_64060[2]	*Streptomyces leeuwenhoekii*	Acetyltransferase	A0A0F7W1L1
Sle_66910[2]	*Streptomyces leeuwenhoekii*	UPF0256 protein SAV_5428	A0A0F7W2A1

[1] Protein sequence translated from DNA in ExPASy Translate Tool (http://web.expasy.org/translate/).
[2] Putative AAC encoded in the genome of *S. leeuwenhoekii* C34.

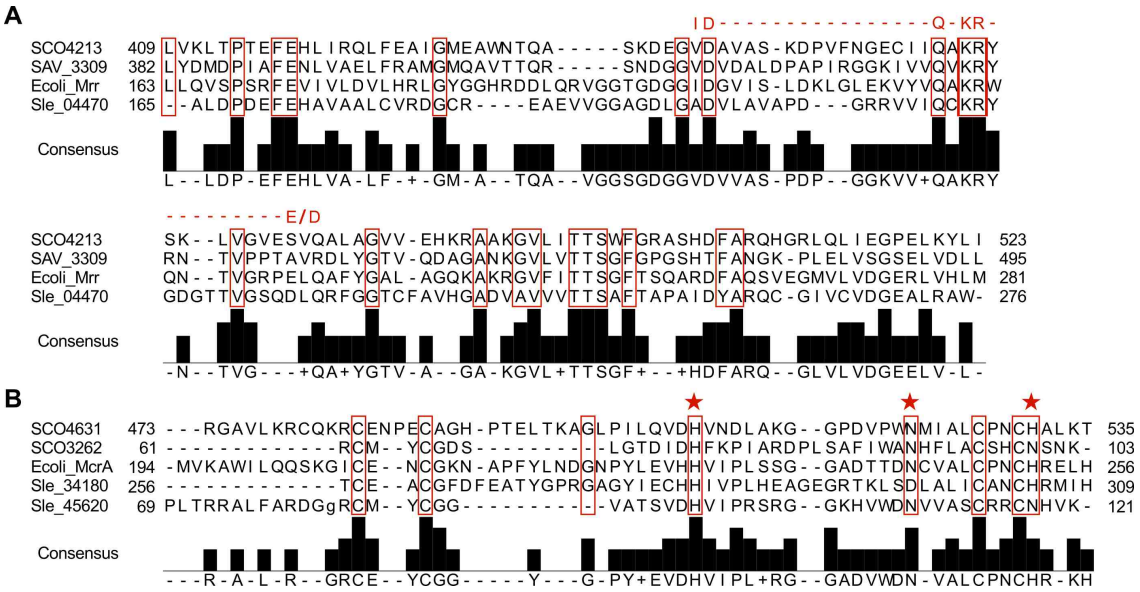


Figure G.5: Alignment of putative proteins of *S. leeuwenhoekii* C34 that encode for a methylation-restriction DNA system. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). Bioinformatic analysis of the *S. leeuwenhoekii* C34 genome revealed the presence of at least 3 putative methylation restriction-systems: Sle_04470, Sle_34180 and Sle_45620 (NCBI accessions: CQR59909, CQR62879 and CQR64020, respectively), homologous to those reported in *S. coelicolor*: Sco4213, Sco4631 and Sco3262 (NCBI accessions: NP_628388, NP_628792 and NP_627474) (González-Cerón et al., 2009). (**A**) Mrr_cat domain containing proteins (Pfam accession: PF04471), found in type II restriction enzymes containing the characteristic F-E-x_{24-42}-I-(**D**/E)-x_{12-15}-(**Q**/E/S/A)-x-**K**-R-x_{6-13}-(**D**/**E**) motif (contained in red boxes) as in Mrr restriction enzyme of *E. coli* K-12 (NCBI accession: NP_418771). (**B**) HNH_c domain containing proteins (CDD accession: CD00085), found in HNH nuclease known to restrict 5-methylcytosine-modified DNA, as McrA of *E. coli* (NCBI accession: NP_415677), HNH possesses the characteristic histidine, asparagine, histidine/asparagine catalytic triad (red stars) and four cysteines, which form a putative zinc finger (cysteines contained in red boxes).

Appendix H

Supplementary information of Chapter Two: Identification and heterologous expression of the chaxamycin biosynthesis gene cluster of *Streptomyces leeuwenhoekii* C34

Table H.1: Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
Sle_10440	693	Alpha-galactosidase (melibiase), highly conserved but no *Streptomyces coelicolor* homolog; *Streptomyces cellulosae* (WP_030676983.1); 599/694 (86)						
Sle_10430	326	Peptidase, highly conserved, 71% identity; ***Streptomyces coelicolor* Sco6773**						
Sle_10420	261	GDSL-like lipase/acylhydrolase, very highly conserved, 92% identity; ***Streptomyces coelicolor* Sco6774**						
Sle10410	201	Carboxymuconolactone decarboxylase, highly conserved but no *Streptomyces coelicolor* homolog; *Streptomyces* sp. NRRL F-4835 (WP_030971727.1); 167/191 (87)						
Sle10400	398	Metallo-beta-lactamase, very highly conserved, 84% identity; ***Streptomyces coelicolor* Sco6776**						
Cxm1	138	Hypothetical protein; *Streptomyces turgidiscabies* (WP_006375853.1); 55/71 (77)						

Continued. Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of S. leeuwenhoekii C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
CxmS	327	Dehydrogenase; *Amycolatopsis mediterranei* (KDO09658.1); 260/327 (80)	RifS - Putative NADH-dependent dehydrogenase (AAS07752.1); 255/322 (79)	Sare1242 - Oxidoreductase domain protein (ABV97147.1); 246/325 (76)				Orf6 - Oxidoreductase-like (AAY46807.1); 36/132 (27)
CxmT	323	Oxidoreductase; *Streptomyces katrae* (WP_030301473.1); 204/326 (63)	RifT - NADH-dependent dehydrogenase (AAC01707.1); 41/265 (53)	Sare1243 - Oxidoreductase domain protein (ABV97148.1); 190/322 (59)				
Cxm4	397	Cytochrome P450; *Actinomadura rifamycini* (WP_026404209.1); 324/397 (82)	Orf0 - Cytochrome P450-like protein (AAC01709.1); 320/397 (81)	Sare1245 - Cytochrome P450 (ABV97150.1); 320/397 (81)	Orf0 - Cytochrome P450-like protein (ADM46355.1); 269/397 (68)	RubF4 - Putative cytochrome P450 (CAI94704.1); 190/408 (47)	Asm30 - Cytochrome P450 (AAM54108.1); 63/202 (31)	GdmP - P450 (PikC subfamily) (AAO06929.1); 149/391 (38)
CxmA	5616	AMP-dependent synthetase and ligase; *Salinispora arenicola* (WP_012181459.1); 3525/4669 (75)	RifA - Rifamycin PKS (AAC01710.1); 1936/2678 (72)	Sare1246 - AMP-dependent synthetase and ligase (ABV97151.1); 3525/4669 (75)	NatA - Polyketide synthase (ADM46356.1); 3425/5078 (67)	RubA - Putative PKS (CAI94682.1); 2207/3714 (59)	AsmD - Polyketide Synthase (AAM54078.1); 1488/2991 (50)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 1395/2703 (52)
CxmB	5363	Rifamycin PKS; *Salinispora arenicola* (WP_019902108.1); 3969/5273 (75)	RifB - Rifamycin PKS (AAC01711.1); 3854/5166 (75)	Sare1247 - Beta-ketoacyl synthase (ABV97152.1); 3969/5273 (75)	NatC - Polyketide synthase (ADM46358.1); 3560/5315 (67)	RubB - Putative PKS (CAI94713.1); 2550/4688 (54)	AsmD - Polyketide Synthase (AAM54078.1); 1208/2218 (54)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 1485/2730 (54)
CxmC	1820	Rifamycin PKS; *Amycolatopsis mediterranei* (WP_013222549.1); 1361/1827 (74)	RifC - Rifamycin PKS (AAC01712.2); 1365/1823 (75)	Sare1248 - Beta-ketoacyl synthase (ABV97153.1); 1329/1844 (72)	NatC - Polyketide synthase (ADM46358.1); 1126/1779 (63)	RubB - Putative PKS (CAI94713.1); 993/1787 (56)	AsmD - Polyketide Synthase (AAM54078.1); 910/1749 (52)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 950/1823 (52)
CxmC	1820	Rifamycin PKS; *Amycolatopsis mediterranei* (WP_013222549.1); 1361/1827 (74)	RifC - Rifamycin PKS (AAC01712.2); 1365/1823 (75)	Sare1248 - Beta-ketoacyl synthase (ABV97153.1); 1329/1844 (72)	NatC - Polyketide synthase (ADM46358.1); 1126/1779 (63)	RubB - Putative PKS (CAI94713.1); 993/1787 (56)	AsmD - Polyketide Synthase (AAM54078.1); 910/1749 (52)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 950/1823 (52)
CxmD	1773	Polyketide synthase; *Streptomyces albogriseolus* (AHD24377.1); 1313/1864 (70)	RifD - Rifamycin PKS (AAC01713.1); 1281/1788 (72)	Sare1249 - Beta-ketoacyl synthase (ABV97154.1); 1252/1817 (69)	NatC - Polyketide synthase (ADM46358.1); 1082/1778 (61)	RubB - Putative PKS (CAI94713.1); 967/1777 (54)	AsmD - Polyketide Synthase (AAM54078.1); 899/1731 (52)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 936/1808 (52)

Continued. Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of S. leeuwenhoekii C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
CxmE	3488	Polyketide synthase; *Salinispora arenicola* (WP_028678212.1); 2468/3522 (70)	RifE - Rifamycin PKS (AAC01714.1); 2434/3497 (70)	Sare1250 - Acyltransferase (ABV97155.1); 2463/3521 (70)	NatC - PKS (ADM46358.1); 2189/3571 (61)	RubB - Putative PKS (CAI94713.1); 1825/3557 (51)	AsmD - Polyketide Synthase (AAM54078.1); 1154/2199 (52)	GdmAII - Polyketide synthase; modules 4-5 (AAO06917.1); 1357/2757 (49)
CxmF	269	*N*-acetyltransferase/amide synthase; *Amycolatopsis vancoresmycina* (WP_004559807.1); 189/257 (74)	RifF - Amide synthase (AAC01715.1); 194/256 (76)	Sare1251 - *N*-acetyltransferase (ABV97156.1); 185/257 (72)	NatF - Amide synthase (ADM46361.1); 158/289 (55)	RubF - Putative amide synthase (CAI94702.1); 125/267 (47)	Asm9 - Amide Synthase (AAM54087.1); 105/258 (41)	GdmF - amide synthase; RifF homolog (AAO06919.1); 107/273 (39)
CxmG	368	Aminodehydroquinate synthase; *Amycolatopsis mediterranei* U32 (YP_003762846.1); 293/347 (84)	RifG - Aminodehydroquinate synthase (AAC01717.1); 290/341 (85)	Sare1253 - 3-dehydroquinate Synthase (ABV97158.1); 284/341 (83)	NatG - 3-dehydroquinate synthase (ADM46363.1); 266/336 (79)	RubG - Putative aDHQ synthase (CAI94724.1); 266/332 (80)	Orf10 - Aminodehydroquinate Synthase (AAC14006.1); 246/335 (73)	Orf2 - Aminodehydroquinate synthase (AAY46803.1); 243/332 (73)
CxmH	406	Phospho-2-dehydro-3-deoxyheptonate aldolase; *Amycolatopsis vancoresmycina* (WP_004559804.1); 311/413 (75)	RifH - AminoDAHP synthase (AAC01718.1); 290/427 (68)	Sare1254 - 3-deoxy-7-phosphoheptulonate synthase (ABV97159.1); 291/391 (74)	NatH - DAHP synthase (ADM46364.1); 272/441 (62)	RubH - Putative aDAHP synthase (CAI94725.1); 295/397 (74)		
CxmI	268	Shikimate dehydrogenase; *Actinomadura rifamycini* (WP_026404009.1); 213/266 (80)	RifI - Aminoquinate dehydrogenase (AAC01719.1); 202/261 (77)	Sare1274 - Shikimate dehydrogenase substrate binding domain protein (ABV97179.1); 144/265 (54)	NatI - Shikimate/quinate dehydrogenase (ADM46365.1); 186/268 (69)	RubI - Putative oxidoreductase (CAI94690.1); 18/44 (41)		
CxmK	386	AHBA synthase; *Amycolatopsis mediterranei* (AAA75105.1); 329/388 (85)	RifK - AHBA synthase (AAC01720.1); 330/388 (85)	Sare1255 - Glutamine−scyllo-inositol transaminase (ABV97160.1); 321/388 (83)	NatK - AHBA synthase (ADM46366.1); 313/384 (82)	RubK - Putative AHBA synthase (CAI94689.1); 293/384 (76)	Orf6 - AHBA synthase (AAC13997.1); 270/365 (74)	Orf5 - AHBA synthase (AAY46806.1); 281/385 (73)
CxmL	358	Oxidoreductase; *Streptomyces* sp. 192 (ADI56541.1); 269/356 (76)	RifL - Putative oxidoreductase (AAS07754.1); 268/357 (75)	Sare1256 - Oxidoreductase domain protein (ABV97161.1); 265/357 (74)	NatL - Oxidoreductase (ADM46367.1); 232/359 (65)	RubL - Putative oxidoreductase (CAI94690.1); 199/345 (58)	Orf7 - Oxidoreductase (AAC14003.1); 203/365 (56)	Orf6 - Oxidoreductase-like (AAY46807.1); 207/362 (57)

Continued. Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of S. leeuwenhoekii C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
CxmM	232	Phosphoglycolate phosphatase; *Amycolatopsis rifamycinica* (KDN20697.1); 191/232 (82)	RifM - Putative phosphatase (AAC01721.1); 193/232 (83)	Sare1257 - AHBA synthesis associated protein (ABV97162.1); 185/232 (80)	NatM - Phosphatase (ADM46368.1); 187/231 (81)	RubM - Putative phosphatase (CAI94691.1); 146/215 (68)	Orf8 - Phosphatase (AAC14004.1); 140/204 (69)	Ahba1b - Phosphatase-like (AAY46808.1); 151/213 (71)
CxmN	307	Kanamycin kinase; *Streptomyces katrae* (WP_030293822.1); 225/295 (76)	RifN - D-kanosamine kinase (AAC01722.2); 164/230 (71)	Sare1275 - ROK family protein (RifN AHBA kinase) (ABV97180.1); 157/290 (54)	NatN - Glucose kinase (ADM46369.1); 194/292 (66)	RubN - Putative kinase (CAI94692.1); 159/298 (53)	Asm22 - Kinase (AAM54100.1); 122/226 (54)	Orf3 - Kinase-like (AAY46804.1); 150/253 (59)
Cxm18	295	*N*5,*N*10-methylene tetrahydromethanopterin reductase; *Amycolatopsis mediterranei* (WP_013222573.1); 204/295 (69)	Orf11 - Putative flavin-dependent oxidoreductase (AAC01735.3); 204/295 (69)			Orf1 - Hypothetical protein (CAI94678.1); 88/269 (33)		
Cxm19	533	3-(3-hydroxyphenyl) propionate hydroxylase; *Streptomyces katrae* (WP_030293823.1); 429/526 (82)	Orf19 - Putative 3-(3-hydroxylphenyl) propionate hydroxylase (AAG52989.1); 388/519 (75)	Sare1268 - Monooxygenase FAD-binding (ABV97173.1); 405/518 (78)	Nat2 - FAD-dependent oxidoreductase (ADM46370.1); 367/540 (68)	RubP1 - Putative hydroxyphenylpropionate (CAI94716.1); 305/522 (58)	Asm11 - Oxygenase (AAM54089.1); 86/273 (32)	GdmM - Similar to Rif19; putative FAD-dependent monooxygenase (AAO06920.1); 279/531 (53)
Cxm20	402	Acyltransferase; *Salinispora arenicola* (WP_029537705.1); 253/407 (62)	Orf20 - Putative polyketide associated protein (AAG52990.1); 252/404 (62)	Sare1262 - Conserved PKS associated protein PapA5 (ABV97167.1); 249/407 (61)				
Cxm21	63	Ferredoxin; *Streptomyces chartreusis* (WP_010040569.1); 43/63 (68)		Sare1261 - Protein of unknown function DUF1271 (SCO7676 ferredoxin) (ABV97166.1); 22/64 (34)				
Cxm22	393	Cytochrome P450; *Streptomyces* sp. NRRL S-118 (WP_031068103.1); 258/399 (65)	Orf4 - Putative cytochrome P450 oxidoreductase (AAC01728.1); 221/403 (55)	Sare1260 - Cytochrome P450 (ABV97165.1); 154/394 (39)	Orf0 - Cytochrome P450-like protein (ADM46355.1); 143/404 (35)	RubF4 - Putative cytochrome P450 (CAI94704.1); 121/410 (30)	Asm30 - Cytochrome P450 (AAM54085.1); 45/186 (24)	GdmP - P450 (PikC subfamily) (AAO06929.1); 177/409 (43)

Continued. Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of S. leeuwenhoekii C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
Cxm24	355	Methyltransferase; *Streptomyces* sp. CNH189 (AGH68906.1); 230/356 (65)	Orf14 - C-27 *O*-methyltransferase (AAC01738.1); 22/87 (25)				Asm7 - Methyltransferase (AAM54085.1); 53/224 (24)	
CxmY	433	Transcriptional regulator; *Streptomyces albogriseolus* (AHD24357.1); 289/430 (67)	Orf36 - Putative regulatory protein (AAS07758.1); 182/416 (44)	Sare1270 - Transcriptional regulator, LuxR family (ABV97175.1); 177/416 (43)	Orf5 - Transcriptional regulator (ADM46353.1); 184/384 (48)	RubRg2 - Putative transcription regulator (CAI94711.1); 113/332 (34)		
CxmZ	246	Hypothetical protein - Response regulators OmpR; *Streptomyces* sp. NRRL B-1347 (WP_030679122.1); 177/242 (73)						
CxmJ	168	3-dehydroquinate dehydratase; *Streptomyces katrae* (WP_030293833.1); 130/147 (88)	RifJ - Aminodehydroquinate dehydratase (AAS07762.1); 133/153 (86)	Sare1277 - 3-dehydroquinate dehydratase (ABV97182.1); 113/140 (81)	NatJ - Aminodehydroquinate dehydratase (ADM46373.1); 104/130 (80)	RubJ - Putative aDHQ dehydratase (CAI94721.1); 108/147 (73)	Asm23 - aminoDHQ Dehydratase (AAM54101.1); 94/127 (74)	Orf4 - aDHQ dehydratase (AAY46805.1); 98/144 (68)
Sle_10090	35	Hypothetical protein *Streptomyces griseus* (WP_030700217.1); 21/31(68)						
Sle_10080	138	Membrane-bound lytic murein transglycosylase D *Thiolapillus brandeum* (WP_041070258.1); 22/54(41)						
Sle_10070	181	Hypothetical protein *Streptomyces sclerotialus* (WP_030622736.1); 52/100(52)						

Continued. Summary of similarity between proteins encoded by the chaxamycin biosynthesis gene cluster of S. leeuwenhoekii C34 and those encoded in other ansamycin-type biosynthesis gene clusters.

CDS	Protein length	Highest identity[1]	Rifamycin (*rif*)[2]	Saliniketal/rifamycin (*sare*)[3]	Naphthomycin (*nat*)[4]	Rubradirin (*rub*)[5]	Ansamitocin (*asm*)[6]	Geldanamycin (*gdm*)[7]
Sle_10060	36	Hypothetical protein *Streptomyces griseus* (WP_037680680.1); 21/24(88)						

This Table summarises the similarity of the proteins encoded in the chaxamycin biosynthesis gene cluster (*cxm*) of *S. leeuwenhoekii* C34 with those present in other ansamycin-type biosynthesis gene clusters. BLASTp (Altschul et al., 1997) analysis was carried out using the NCBI non-redundant database or using specific coding protein sequences of ansamycin-type biosynthesis gene clusters to infer possible chaxamycin biosynthesis reactions by comparison with known biosynthesis pathways. The information given in each cell of the Table is: annotated function, species and its accession number in parentheses; identical residues/total residues considered in the alignment (%identity).

[1] Protein found with the highest similarity to that of the *cxm* in the non-redundant database of NCBI (accessed on 13 August 2014).

[2] Rifamycin biosynthesis gene cluster (*rif*) from *Amycolatopsis. mediterranei* S699 (AF040570.3).

[3] Saliniketal/rifamycin biosynthesis gene cluster (*sare*) from *Salinispora arenicola* strain CNS-205 (CP000850).

[4] Naphthomycin biosynthesis gene cluster (*nat*) from *Streptomyces* sp. CS (GQ452266.1)

[5] Rubradirin biosynthesis gene cluster (*rub*) from *Streptomyces achromogenes* subsp. *rubradiris* (AJ871581.1).

[6] Ansamitocin biosynthesis gene cluster (*asm*) from *Actinosynnema pretiosum* subsp. *auranticum* (AF453501.1 and U33059.1).

[7] Geldanamycin biosynthesis gene cluster (*gdm*) from *Streptomyces geldanamycininus* (AY952143 and AY179507).

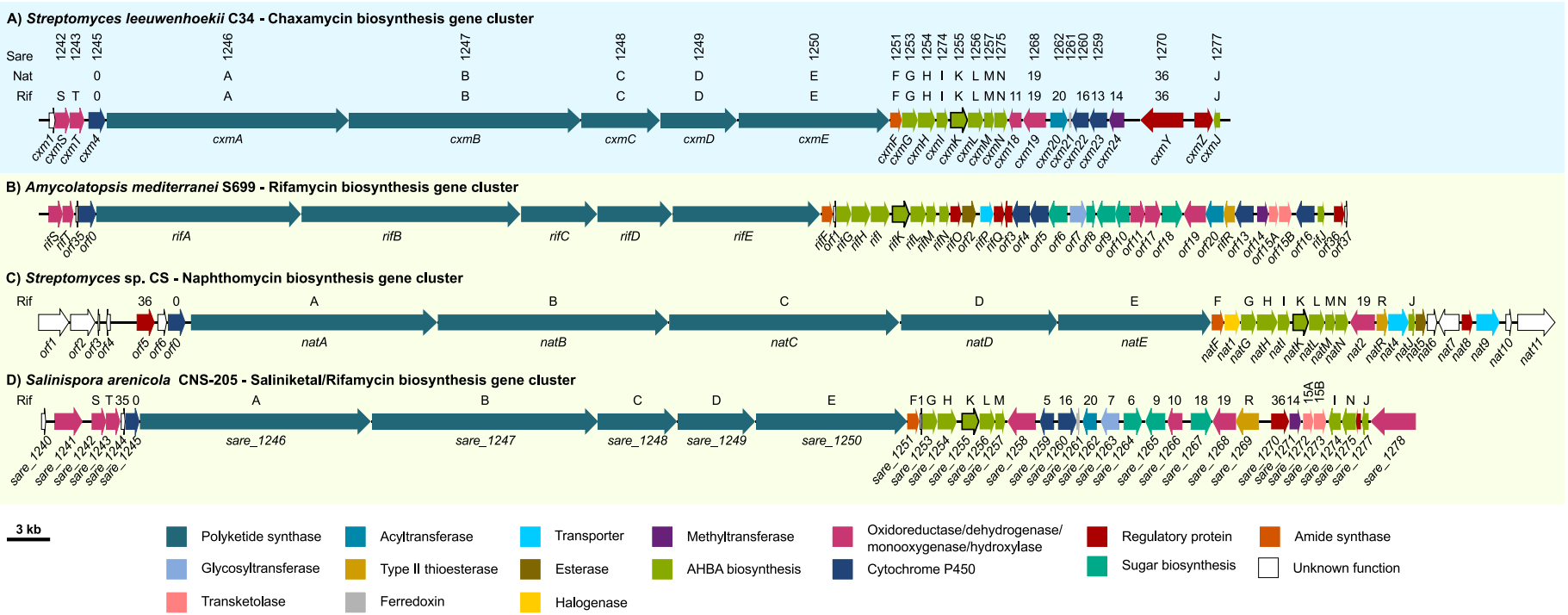


Figure H.1: Comparison of the chaxamycin biosynthesis gene cluster (*cxm*) of *S. leeuwenhoekii* C34 with other ansamycin-type biosynthesis gene clusters. (**A**) Chaxamycin biosynthesis gene cluster (*cxm*); (**B**) saliniketal/rifamycin biosynthesis gene cluster (*sare*) from *Salinispora arenicola* CNS-205 (CP000850.1); (**C**) rifamycin biosynthesis gene cluster (*rif*) from *Amycolatopsis mediterranei* S699 (AF040570.3) and (**D**) naphthomycin biosynthesis gene cluster (*nat*) from *Streptomyces* sp. CS (GQ452266.1). Annotations on the top of some genes indicate that that gene shares high homology with genes present in other ansamycin-type biosynthesis gene cluster (based on Table Appendix H.1). Colour code for genes and a scale bar are shown in the bottom of this Figure.

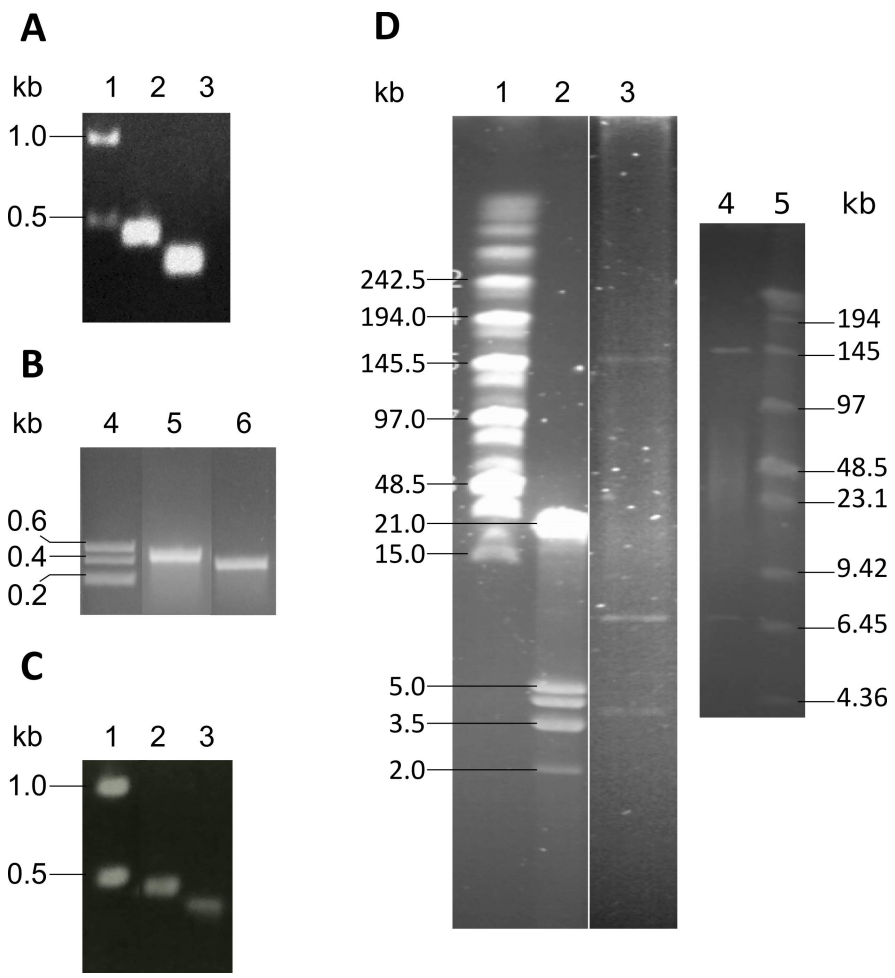


Figure H.2: Screening for the chaxamycin biosynthesis gene cluster in a PAC genomic library of *S. leeuwenhoekii* C34. (**A**) Amplification by PCR of the predicted ends of the chaxamycin biosynthesis gene cluster using genomic DNA of *S. leeuwenhoekii* C34: lane 1, 1 kb DNA ladder (cat. no. N3232S; NEB, USA); lane 2, PCR with primers JFC022/JFC023 that amplify a fragment of 439 bp, representing the right end of the cluster (nt position 1,292,956 to 1,293,394); lane 3, PCR with primers JFC024/JFC025 that amplify a fragment of 340 bp, representing the left end of the cluster (nt position 1,211,083 to 1,211,422); sequences were verified by Sanger sequencing. (**B**) Result of the screening for the chaxamycin biosynthesis gene cluster in a PAC genomic library of *S. leeuwenhoekii* C34, carried out by Bio S&T (Montreal, Canada). The result corresponds to the clone 2-11L (pIJ12853) that was positive for PCR with both pair of primers mentioned in **A**: lane 1: DNA ladder; lane 2, PCR with primers JFC022/JFC023 (expected amplicon size 439 bp); lane 3, PCR with primers JFC024/JFC025 (expected amplicon size 340 bp). (**C**) PCR verification of the PAC clone 2-11L (pIJ12853) received from Bio S&T, that harbours the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34: lane 4, 1 kb DNA ladder (cat. no. N3232S; NEB, USA); lane 5, PCR with primers JFC022/JFC023 (expected amplicon size 439 bp); lane 5, PCR with primers JFC024/JFC025 (expected amplicon size 340 bp). (**D**) Pulse field electrophoresis gel of the digestion of PAC clone 2-11L (pIJ12853) with *Dra*I restriction enzyme in 1% agarose gel: lane 1, high molecular weight DNA ladder (cat. no. N3551S; NEB, USA); lane 2, lambda DNA digested with *Eco*RI/*Hind*III (cat. no. G173A; Promega, USA); lane 3, result of the digestion of PAC clone 2-11L (pIJ12853) received from Bio S&T; lane 4, control result of the digestion of PAC clone 2-11L (pIJ12853), carried out by Bio S&T; lane 5, DNA ladder.

Table H.2: BLASTp analysis of the coding sequences found in the *S. leeuwenhoekii* C34 genomic insert present in pIJ12853.

Sle_number	NCBI accession	Proposed function	Homologue (% identity; UniRef90 accession)
Sle_09610 (partial: 389/511 aa)	CQR60424	Na(+)/H(+) antiporter	*Streptomyces* sp. LaPpAH-108 (34.6%; UPI00035CE13C)
Sle_09620	CQR60425	dTDP-glucose 4,6-dehydratase	*Micromonospora* sp. L5 (64.8%; D9TDA3)
Sle_09630	CQR60426	NDP-Hexose 2,3-Dehydratase	*Streptomyces acidiscabies* (50.3%; UPI000288F3A1)
Sle_09640	CQR60427	Hypothetical protein	*Streptomyces violaceoruber* (29.9%; C0Z479)
Sle_09650	CQR60428	Transposase	*Streptomyces ipomoeae* (88.0%; N0D0W7)
Sle_09660	CQR60429	Transposase	*Streptomyces ambofaciens* ATCC 23877 (89.1%; Q1RQM2)
Sle_09670	CQR60430	Hypothetical protein	*Streptomyces ambofaciens* ATCC 23877 (42.2%; UPI0003615AD0)
Sle_09680	CQR60431	Transposase	*Streptomyces ambofaciens* ATCC 23877 (84.5%; L1KIU3)
Sle_09690	CQR60432	Phage integrase	*Streptomyces coelicolor* A3(2) (85.8%; A0A0F7VSG2)
Sle_09700	CQR60433	Transketolase A/Ferredoxin Fas2	*Streptomyces ghanaensis* ATCC 14672 (89.4%; A0A0C5GC69)
Sle_09710	CQR60434	Transketolase B/Uncharacterized 33.6 kDa protein	*Streptomyces viridochromogenes* (87.1%; D9X8L6)
Sle_09720	CQR60435	Hypothetical Protein	*Streptomyces* sp. Amel2xE9 (34.5%; UPI0003702C97)
Sle_09730	CQR60436	Conserved hypothetical protein	*Streptomyces* (91.7%; L8PPY7)
Sle_09740	CQR60437	Acetyltransferase	*Streptomyces* sp. 303MFCol5.2 (%; UPI000363BB59)
Sle_09750	CQR60438	XRE family transcriptional regulator	*Streptomyces lividans* TK24 (86.6%; D6EF12)
Sle_09760	CQR60439	HTH-type transcriptional regulator gltC	*Streptomyces* (91.7%; V6KT60)
Sle_09770	CQR60440	Putative 8-amino-7-oxononanoate synthase/2-amino-3-ketobutyrate coenzyme A ligase	*Streptomyces scabiei* (strain 87.22) (94.5%; C9Z6N6)
Sle_09780	CQR60441	L-threonine 3-dehydrogenase	*Streptomyces coelicolor* A3(2) (90.7%; Q9L233)
Sle_09790	CQR60442	Conserved hypothetical protein	*Streptomyces davawensis* JCM 4913 (72.5%; K4QYF7)
Sle_09800	CQR60443	Hypothetical	*Streptomyces lividans* TK24 (86.4%; A0A076LYX9)
Sle_09810	CQR60444	Conserved hypothetical protein	*Streptomyces* sp. LaPpAH-108 (87.9%; UPI00036635BC)

Continued. BLASTp analysis of the coding sequences found in the S. leeuwenhoekii C34 genomic insert present in pIJ12853.

Sle_number	NCBI accession	Proposed function	Homologue (% identity; UniRef90 accession)
Sle_09820	CQR60445	Roadblock/LC7 family protein	*Streptomyces* (86.1%; Q82MM8)
Sle_09830	CQR60446	Histidine kinase	*Streptomyces davawensis* JCM 4913 (71.2%; K4QT85)
Sle_09840	CQR60447	Hypothetical protein	*Actinoplanes* sp. N902-109 (40.7%; R4LU19)
Sle_09850	CQR60448	Wall-associated protein	*Streptomyces* sp. CNR698 (72.9%; UPI0004B1F034)
Sle_09860	CQR60449	Conserved hypothetical protein	*Streptomyces cyaneogriseus* (80.3%; UPI00069BC254)
Sle_09870	CQR60450	Hypothetical protein	*Streptomyces viridochromogenes* Tue57 (65.2%; L8P6I9)
Sle_09880	CQR60451	Hypothetical protein	*Streptomyces coelicoflavus* ZG0656 (71.9%; H1Q8T4)
Sle_09890	CQR60452	Uncharacterized HTH-type transcriptional regulator PA4778	*Streptomyces viridochromogenes* Tue57 (82.4%; L8P667)
Sle_09900	CQR60453	Conserved hypothetical protein	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* strain NMWT 1 (84.9%; A0A0C5G9B6)
Sle_09910	CQR60454	Restriction Endonuclease	*Streptomyces* sp. Tu 6176 (89.5%; A0A022M843)
Sle_09920	CQR60455	GDSL-like protein	*Streptomyces ipomoeae* 91-03 (73.4%; L1KM30)
Sle_09930	CQR60456	HTH-type transcriptional regulator CelR	*Streptomyces regensis* (91.0%; A0A0J8AJB3)
Sle_09940	CQR60457	Endoglucanase 1	*Streptomyces himastatinicus* ATCC 53653 (92%; A0A0F7VSJ8)
Sle_09950	CQR60458	HTH-type transcriptional regulator glxA	*Streptomyces viridochromogenes* Tue57 (87.7%; L8PHZ6)
Sle_09960	CQR60459	ThiJ/PfpI domain-containing protein	*Streptomyces viridochromogenes* Tue57 (86.8%; L8PP34)
Sle_09970	CQR60460	ThiJ/PfpI domain-containing protein	*Streptomyces viridochromogenes* Tue57 (89.0%; L8PMH1)
Sle_09980	CQR60461	Endo-1,4-beta-xylanase B	*Streptomyces* sp. 303MFCol5.2 (50.3%; UPI00036670A5)
Sle_09990	CQR60462	Hypothetical C	*Streptomyces collinus* Tu 365 (30.9%; S5W1H4)
Sle_10000	CQR60463	Conserved hypothetical protein	*Streptomyces* sp. Mg1 (50.0%; B4V7A7)
Sle_10010	CQR60464	Conserved hypothetical protein	*Streptomyces* sp. NRRL S-146 (93.0%; A0A0F7VLK2)
Sle_10020	CQR60465	von Willebrand factor type A	*Streptomyces* (88.9%; UPI0004BE9351)

Continued. BLASTp analysis of the coding sequences found in the S. leeuwenhoekii C34 genomic insert present in pIJ12853.

Sle_number	NCBI accession	Proposed function	Homologue (% identity; UniRef90 accession)
Sle_10030	CQR60466	Conserved hypothetical protein	*Streptomyces cyaneogriseus* (96.0%; A0A0F7VU26)
Sle_10040	CQR60467	Hypothetical protein	*Succinimonas amylolytica* (16.5%; UPI00036F57F6)
Sle_10050	CQR60468	Hypothetical protein	no-hit in UniRef90
Sle_10060	CQR60469	Hypothetical protein	*Streptomyces tendae* (59.5%; A7DWM0)
Sle_10070	CQR60470	Hypothetical protein LILAB	*Kitasatospora setae* (strain ATCC 33774) (37.4%; E4N4F8)
Sle_10080	CQR60471	Hypothetical protein	*Haemonchus contortus* (31.9%; U6NVP5)
Sle_10090	CQR60472	Hypothetical protein	*Streptomyces* (61.1%; Q1HVZ8)
CxmJ	CQR60473	RifJ homologue, AminoDHQ dehydratase, AHBA biosynthesis	*Amycolatopsis vancoresmycina* DSM 44592 (74.0%; R1GB14)
CxmZ	CQR60474	Transcriptional regulator, atypical response regulator	*Streptomyces bingchenggensis* (strain BCW-1) (44.5%; D7BUX8)
CxmY	CQR60475	LuxR Family transcriptional regulator, possible activator; no Rif homologue	*Streptomyces griseochromogenes* (46.1%; C6ZCQ9)
CxmR2	CQR60476	RifR homologue; Type II thioesterase, probably non-functional	Uncultured bacterium esnapd26 (51.9%; S5UC25)
CxmR	CQR60477	RifR homologue; Type II thioesterase, probably non-functional	*Streptomyces rochei* (69.8%; Q83X18)
Cxm24	CQR60478	Methyltransferase SAM-dependent; not homolog to rifamycin Orf14	*Streptomyces vitaminophilus* (69.1%; UPI00037DA427)
Cxm23	CQR60479	Rif-orf5 homologue; Cytochrome P450	*Micromonosporaceae* (76.8%; A8M6D9)
Cxm22	CQR60480	Rif-orf4 homologue; Cytochrome P450	*Streptomyces* sp. Amel2xE9 (65.5%; UPI000369715D)
Cxm21	CQR60481	Ferredoxin 4Fe-4S single cluster domain; no Rif homologue	*Streptomyces hygroscopicus* subsp. *jinggangensis* (64.1%; H2JZE7)
Cxm20	CQR60482	Rif-orf20 homologue; *O*-Acyltransferase	*Salinispora arenicola* (strain CNS-205) (62.0%; A8M6E2)
Cxm19	CQR60483	Rif-orf19 homologue. FDA dependent-monooxygenase; Naphthalene ring formation; 3-(3-hydroxy-phenyl)propionate/3-hydroxycinnamic acid hydroxylase	*Salinispora arenicola* (strain CNS-205) (78.0%; A8M6Z1)
Cxm18	CQR60484	Rif-orf11 homologue, flavin-dependent oxidoreductase	*Amycolatopsis mediterranei* (S699) (69.2%; G0FUR3)

Continued. BLASTp analysis of the coding sequences found in the S. leeuwenhoekii C34 genomic insert present in pIJ12853.

Sle_number	NCBI accession	Proposed function	Homologue (% identity; UniRef90 accession)
CxmN		RifN homologue, D-kanosamine kinase, AHBA biosynthesis	*Amycolatopsis mediterranei* (strain S699) (64.6%; G0FS68)
CxmM	CQR60486	RifM homologue, phosphoglycolate phosphatase, AHBA biosynthesis	*Amycolatopsis vancoresmycina* DSM 44592 (85.0%; R1IGA2)
CxmL	CQR60487	RifL homologue, oxidoreductase, AHBA biosynthesis	*Amycolatopsis mediterranei* (strain S699) (74.9%; Q7BUE1)
CxmK	CQR60488	RifK homologue, 3-amino-5-hydroxy benzoic acid (AHBA) synthase	*Amycolatopsis mediterranei* (strain S699) (85.8%; O52552)
CxmI	CQR60489	RifI homologue, Aminoquinate dehydrogenase, AHBA biosynthesis	*Actinomadura rifamycini* (86.0%; UPI000402CE84)
CxmH	CQR60490	RifH homologue, AminoDAHP synthase, AHBA biosynthesis	*Amycolatopsis vancoresmycina* DSM 44592 (76.4%; R1HT87)
CxmG	CQR60491	RifG homologue, AminoDHQ synthase, AHBA biosynthesis	*Amycolatopsis ryfamicinica* (85.7%; A0A066U4E0)
CxmF	CQR60492	RifF homologue, amide synthase, release of polyketide from PKS and cyclisation	*Amycolatopsis mediterranei* (strain S699) (71.5%; O52547)
CxmE	CQR60493	Polyketide synthase, type I, modules: 9 and 10	*Salinispora arenicola* (strain CNS-205) (70.6%; A8M6D0)
CxmD	CQR60494	Polyketide synthase, type I, module 8	*Streptomyces albogriseolus* (70.5%; V9XSJ9)
CxmC	CQR60495	Polyketide synthase, type I, module 7	*Salinispora arenicola* (74.5%; UPI000524BDFB)
CxmB	CQR60496	Polyketide synthase, type I, modules: 4, 5 and 6	*Salinispora arenicola* CNS-205 (75.4%; A8M6C7)
CxmA	CQR60497	Polyketide synthase, type I, modules: loading, 1, 2 and 3	*Streptomyces* sp. CS (66.8%; F8QPF8)
Cxm4	CQR60498	Rif-orf0(cero) homologue, Cytochrome P450	*Actinomadura rifamycini* (81.6%; UPI00047DA12C)
CxmT	CQR60499	RifT homologue; NAD-dependent oxydoreductase	*Salinispora arenicola* (strain CNS-205) (58.6%; A8M6C3)
CxmS	CQR60500	RifS homologue; NAD-dependent oxydoreductase	*Amycolatopsis mediterranei* (*Nocardia mediterranei*) (79.2%; Q7BUE3)
Cxm1 (Sle_10390)	CQR60501	Hypothetical protein	*Streptomyces turgidiscabies* Car8 (40.3%; L7FBL5)
Sle_10400	CQR60502	Metallo-beta-lactamase, very highly conserved, 84% identity Sco6776	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* strain NMWT 1 (95.0%; A0A0F7VSN7)
Sle_10410	CQR60503	Carboxymuconolactone decarboxylase; highly conserved but no Sco homologue	*Streptomyces cyaneogriseus* (98.0%; UPI0005C98431)

Continued. BLASTp analysis of the coding sequences found in the S. leeuwenhoekii C34 genomic insert present in pIJ12853.

Sle_number	NCBI accession	Proposed function	Homologue (% identity; UniRef90 accession)
Sle_10420	CQR60504	GDSL-like Lipase/Acylhydrolase, very highly conserved, 92% identity Sco6774	*Streptomyces* sp. LaPpAH-108 (86.6%; UPI00037D495F)
Sle_10430	CQR60505	Peptidase, highly conserved, 71% identity Sco6773	*Streptomyces ambofaciens* (72.1%; UPI00069DC1A9)
Sle_10440	CQR60506	Alpha-galactosidase (melibiase), highly conserved but no Sco homologue	*Streptomyces turgidiscabies* Car8 (85.0%; L7F8E0)
Sle_10450	CQR60507	Putative tyrosine-protein phosphatase H16_A0669; 89% identity Sco6772	*Streptomyces* (88.0%; Q9X7W7)
Sle_10460	CQR60508	Possible small hydrophobic secreted protein; 82% identity Sco6771 (BASys: Conserved Hypothetical Protein)	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* strain NMWT 1 (100%; K4R009)
Sle_10470	CQR60509	XRE Family Transcriptional Regulator; 84% identity Sco6770	*Streptomyces* sp. PsTaAH-124 (84.5%; UPI00036204E7)
Sle_10480	CQR60510	Acetylornithine aminotransferase; 89% identity Sco6769	*Streptomyces* (90.5%; D6A2X5)
Sle_10490	CQR60511	1-deoxy-D-xylulose-5-phosphate synthase 1; 90% identity Sco6768; 73% identity Sco6013	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* (98.0%; A0A0F7VU62)
Sle_10500	CQR60512	Molybdenum cofactor biosynthesis protein A 1; 94% identity Sco6766/Hopanoid biosynthesis associated radical SAM protein HpnH	*Streptomyces turgidiscabies* Car8 (90.9%; L7F9E9)
Sle_10510	CQR60513	Lipoprotein; 84% identity Sco6765	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* strain NMWT 1 (95.0%; A0A0F7VPP4)
Sle_10520 (partial: 533/675 aa)	CQR60514	Squalene-hopene cyclase; 88% identity Sco6764	*Streptomyces cyaneogriseus* subsp. *noncyanogenus* strain NMWT 1 (99.0%; A0A0F7VLM6)

This Table summarises BLASTp analysis of the proteins encoded in the 145 kb genomic insert cloned into pIJ12853. This insert contains the putative chaxamycin biosynthesis gene cluster (80.2 kb) of *S. leeuwenhoekii* C34. Local BLASTp (Altschul et al., 1997; Priyam et al.) analysis was carried out using the Uniref90 database of UniProt (http://www.uniprot.org/help/uniref; Suzek et al., 2007) (analysis performed on 23 March 2014). Genes marked in red encode for putative transketolases subunits; genes marked in blue indicate the chaxamycin biosynthesis gene cluster; genes marked in orange indicate homologues to the hopanoid biosynthesis gene cluster.

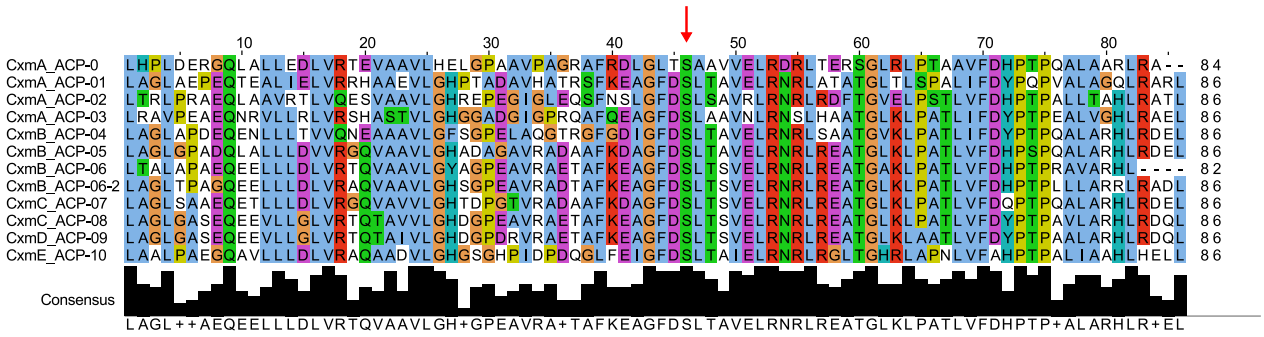


Figure H.3: Alignment of amino acid sequences of ACP domains of the chaxamycin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that ACP is located in the chaxamycin PKS ("CxmA_ACP-0", refers to the ACP of the Loading Module of the PKS subunit CxmA; "CxmA_ACP-01", refers to the ACP of Module 1 of the PKS subunit CxmA, and so on). The arrow indicates the position of the conserved serine residue that harbours the 4'-phosphopantetheine prosthetic group, which is a diagnostic residue to infer functionality of this domain.

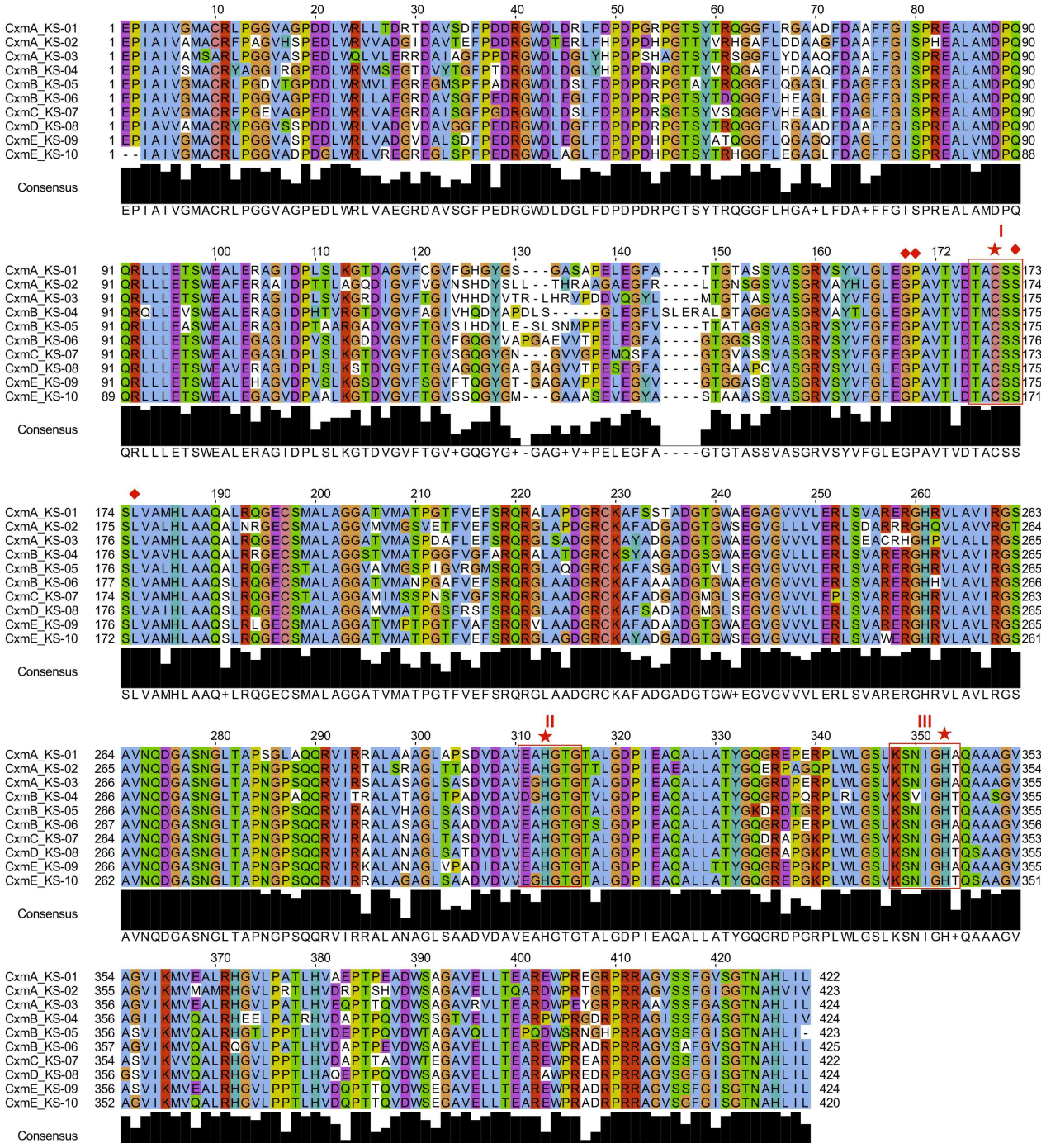


Figure H.4: Alignment of amino acid sequences of KS domains of the chaxamycin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that KS is located in the chaxamycin PKS ("CxmA_KS-01", refers to the KS of Module 1 of the PKS subunit CxmA; "CxmA_KS-02", refers to the KS of Module 2 of the PKS subunit CxmA, and so on). The residues of the conserved motif GPxxxxxxC(S/A/T)(S/T)x(L/V) are indicated with a red diamond (Fernández-Moreno et al., 1992); catalytic triad cysteine, histidine and histidine present in motifs I, II and III, respectively are indicated with red stars (Keatinge-Clay, 2012).

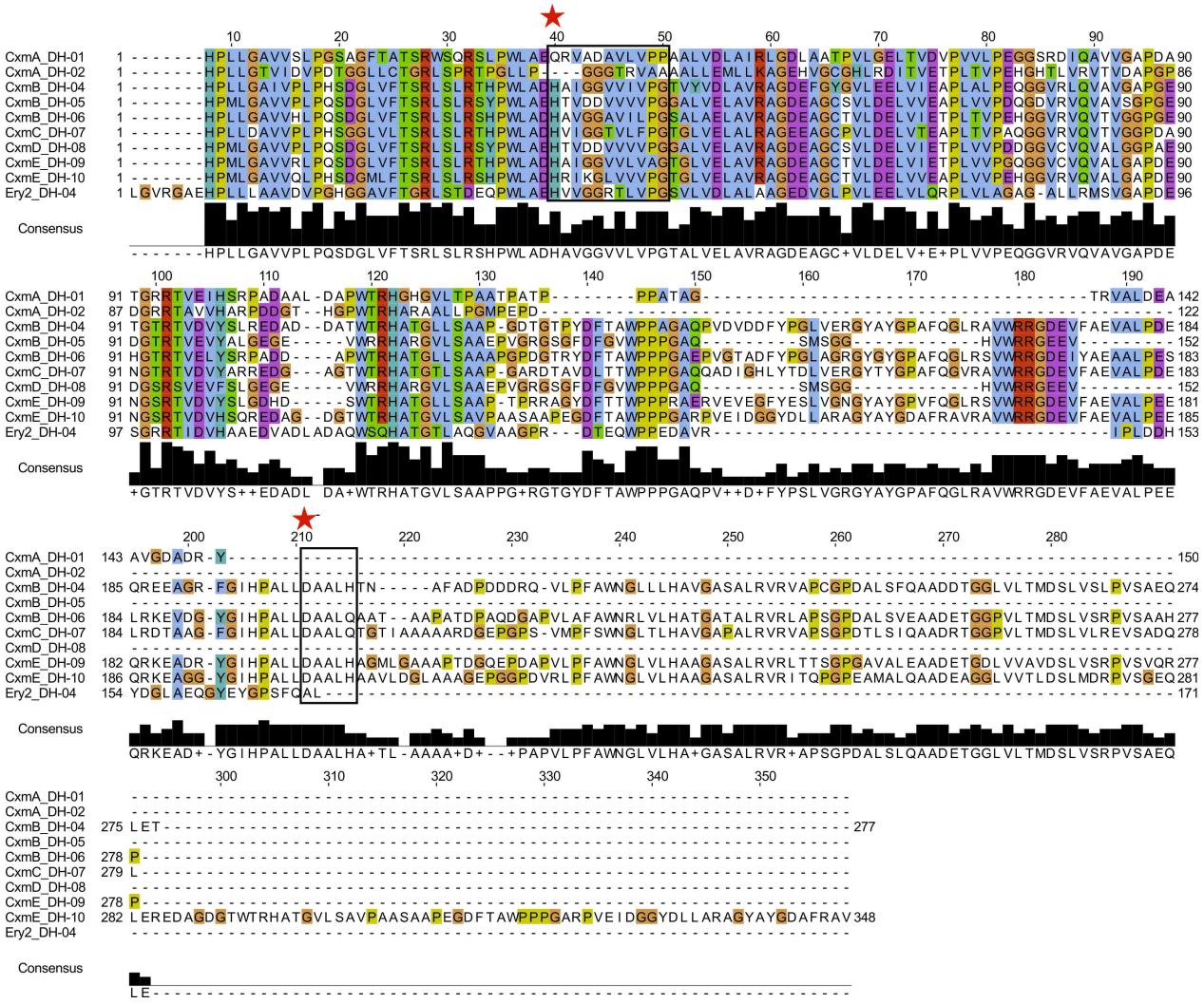


Figure H.5: Alignment of amino acid sequences of DH domains of the chaxamycin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that DH is located in the chaxamycin PKS ("CxmA_DH-01", refers to the DH of Module 1 of the PKS subunit CxmA; "CxmA_DH-02", refers to the DH of Module 2 of the PKS subunit CxmA and so on). The conserved motifs are shown contained in black boxes and residues histidine and aspartate of the catalytic dyad are indicated with red stars.

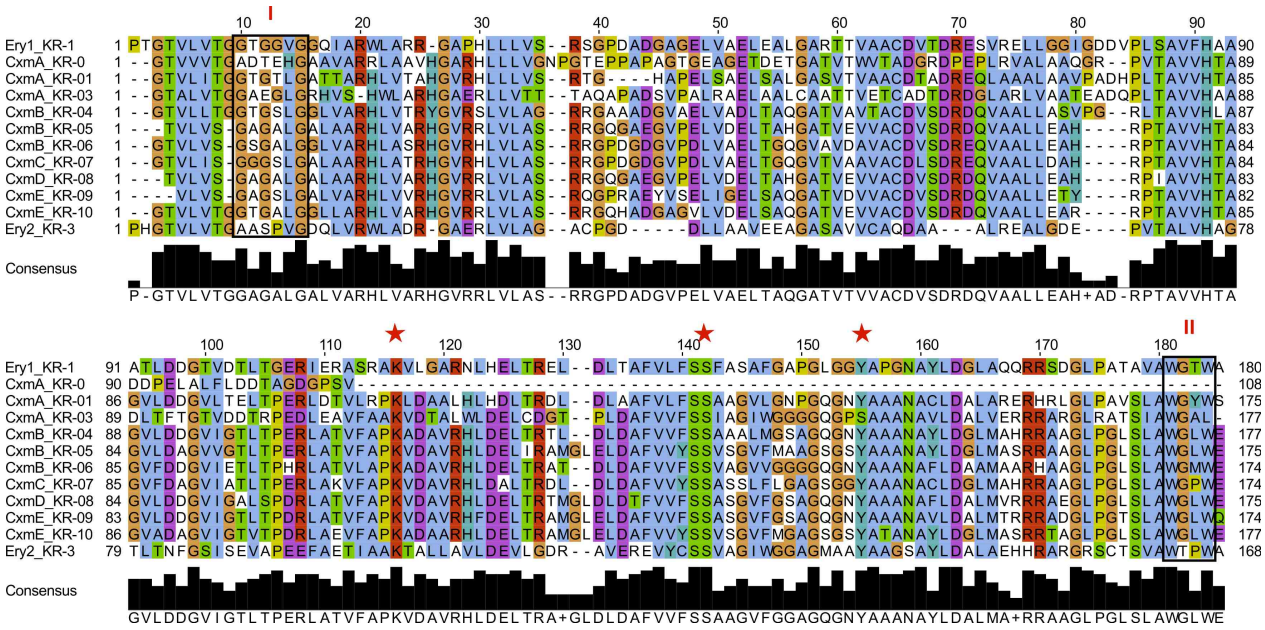


Figure H.6: Alignment of amino acid sequences of KR domains of the chaxamycin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that KR is located in the chaxamycin PKS ("CxmA_KR-0", refers to the KR of Loading Module of the PKS subunit CxmA; "CxmA_KR-01", refers to the KR of Module 1 of the PKS subunit CxmA, and so on). Reductase-competent KR domains contain the conserved motifs that allow NADPH binding, GxGxxG (I) and WGxW (II), which are contained in black boxes. Erythromycin KR domains 1 (active) and 3 (inactive) are shown for reference. The catalytic triad in the active site, lysine, serine and tyrosine, are indicated with red stars, as in Reid et al., 2003.

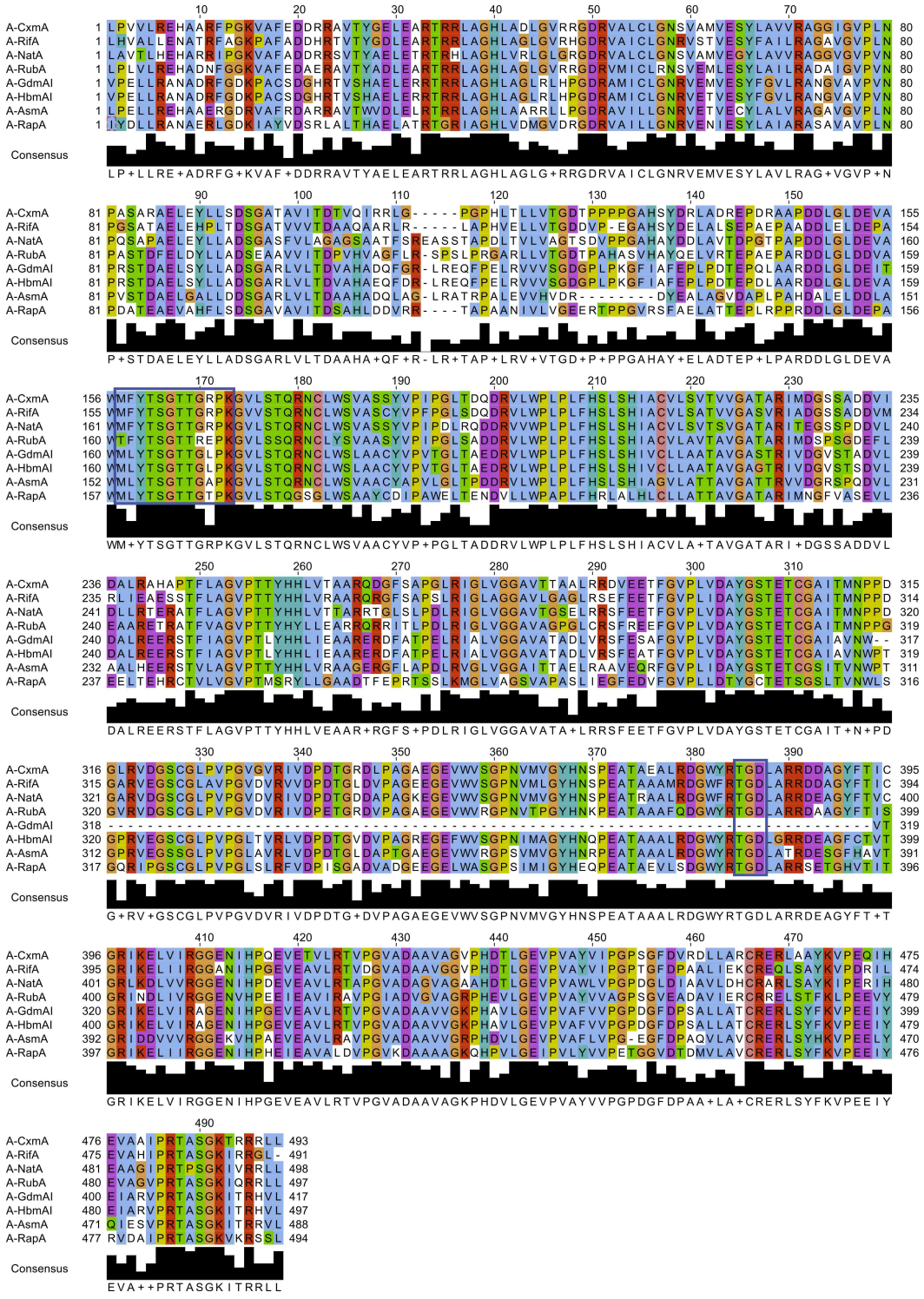


Figure H.7: Comparison of the amino acid sequence of the A domain present in the Loading Module of the chaxamycin PKS with other adenylation domains from PKSs of known specificity. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The AMP-binding [LIVMFY]-{E}-{VES}-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-x-[PASLIVM]-[KR] motif (PROSITE accession number: PS00455; Sigrist et al., 2012) and the conserved ATPase TGD motif (Huang et al., 2001) are shown in blue boxes. See nomenclature of the proteins on the *next page*.

Nomenclature for proteins used in the Figure H.7

Adenylation domain	Identity %[1]	biosynthesis gene cluster	Substrate specificity	Microorganism
A-CxmA	-	Chaxamycin	AHBA	*Streptomyces leeuwenhoekii* C34
A-RifA	77	Rifamycin	AHBA	*Amycolatopsis mediterranei* S699
A-NatA	76	Naphthomycin	AHBA	*Streptomyces* sp. CS
A-RubA	55	Rubradirin	AHBA	*Streptomyces achromogenes* var. *rubradiris* NRRL 3061
A-GdmAI	61	Geldanamycin	AHBA	*Streptomyces geldanamycininus* NRRL 3602
A-HbmAI	67	Herbimycin	AHBA	*Streptomyces hygroscopicus* AM 3672
A-AsmA	63	Ansamitocin	AHBA	*Actinosynnema pretiosum* subsp. *auranticum* ATCC 31565
A-RapA	55	Rapamycin	DHCHC	*Streptomyces rapamycinicus* NRRL 5491

[1] Identity percentage respect to the adenylation domain of CxmA.

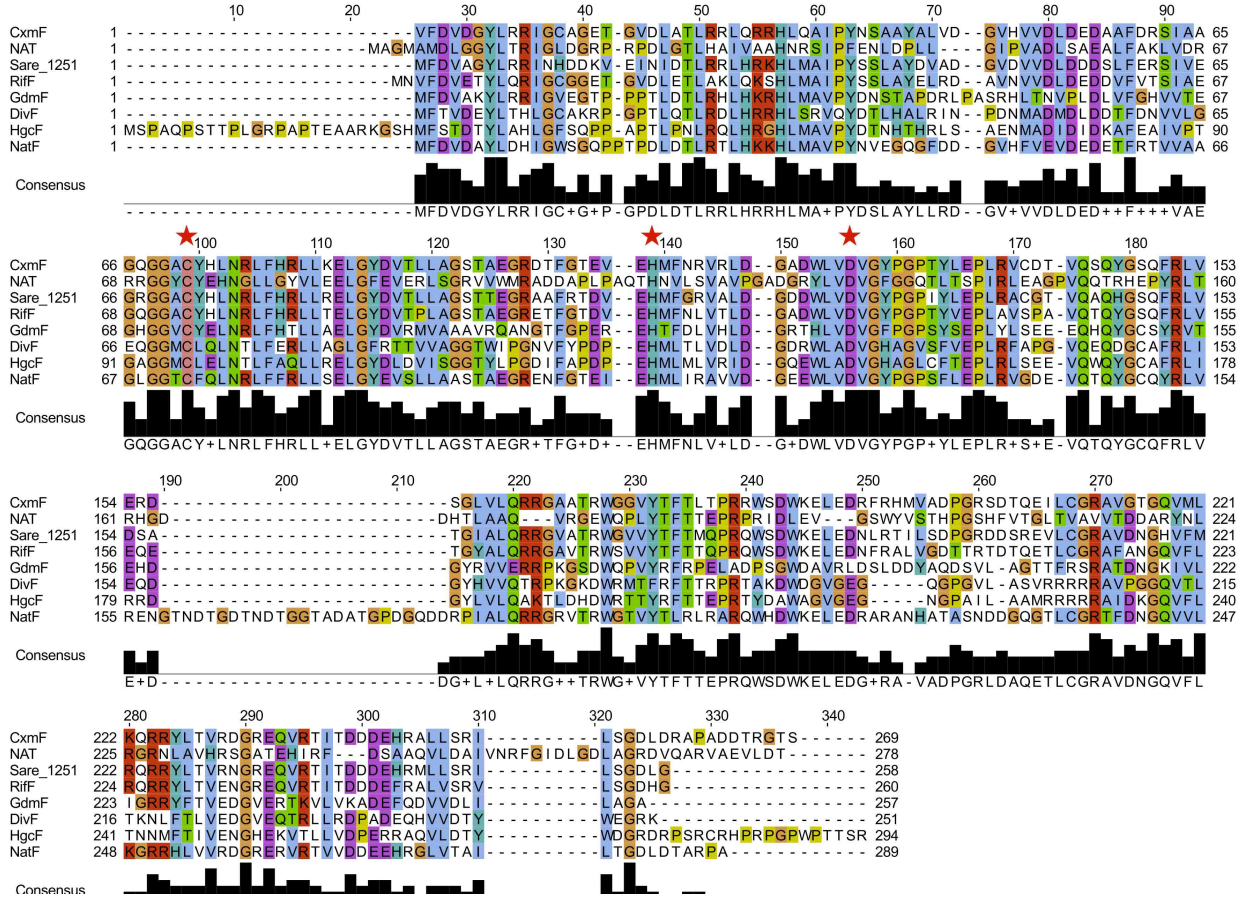


Figure H.8: Alignment of amino acid sequences of amide synthase enzyme (CxmF) found in the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 with others known amide synthases. The amino acid sequences were obtained from the NCBI website, aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The conserved residues cysteine, histidine and aspartate of the catalytic triad are indicated with red stars. NCBI accession in parentheses. CxmF of *S. leeuwenhoekii* C34 (CQR60492.1); NAT (arylamine *N*-acetyltransferase) of *Mycobacterium smegmatis* were used as reference as in Sim et al., 2008 (CKG90611.1). Amide synthases from other ansamycin-type biosynthesis gene clusters like, Sare_1251 of *Sal. arenicola* CNS-205 (ABV97156.1); RifF of *A. mediterranei* S699 (AAC01715.1); GdmF of *S. hygroscopicus* (AAO06919.1); DivF of *Streptomyces* sp. W112 (AHG95681); HgcF of *Streptomyces* sp. LZ35 (AHG95681); NatF of *Streptomyces* sp. CS (ADM46361) were also used as reference.

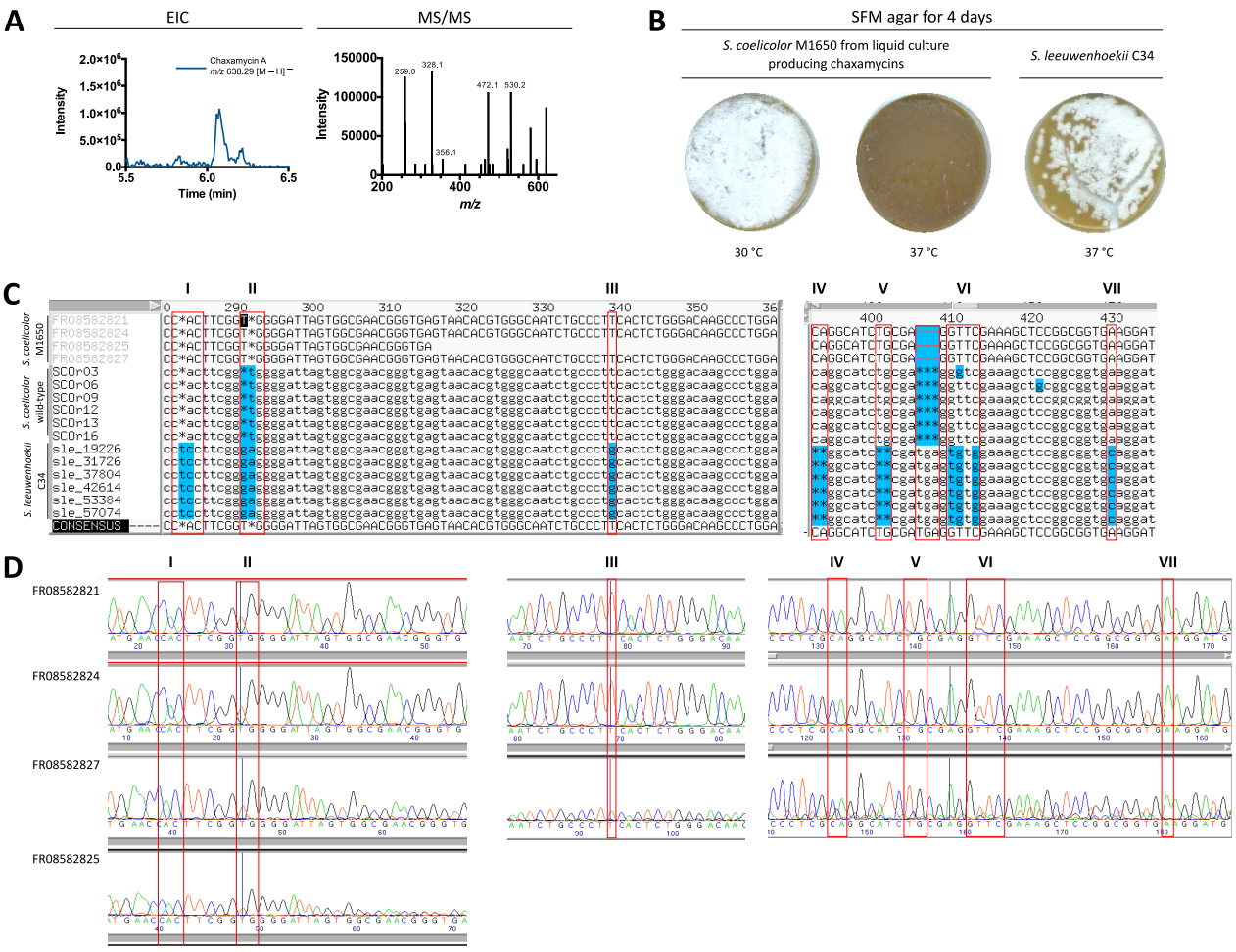


Figure H.9: Assessment of the purity of the culture of *S. coelicolor* M1650 that produces chaxamycins and verification of the 16S rRNA gene sequence. (**A**) EIC and MS/MS fragmentation of chaxamycin A *m/z* 638.29 $[M - H]^-$ detected in the supernatant of a culture of *S. coelicolor* M1650 (*S. coelicolor* M1152 carrying pIJ12853 that contains the chaxamycin biosynthesis gene cluster). (**B**) Two aliquots of the culture of *S. coelicolor* M1650 were streaked out on SFM agar and one was incubated at 30 °C and the second one was cultivated at 37 °C. At 37 °C *S. coelicolor* does not grow well. In addition, *S. leeuwenhoekii* C34 was streaked on SFM agar and incubated at 37 °C as control. The result indicates that the culture of *S. coelicolor* M1650 was pure since growth of *S. coelicolor* M1650 was observed at 30 °C and no *S. leeuwenhoekii* C34 colonies were observed on the plate of *S. coelicolor* M1650 incubated at 37 °C. (**C**) Amplification of 16S rRNA gene from genomic DNA of *S. coelicolor* M1650 isolated from the culture producing chaxamycins using primers 27F and 1492R, and alignment of the amplified sequences (FR08582821, FR08582824, FR08582825 and FR08582827) against 16S rRNA genes of *S. leeuwenhoekii* C34 (*sle_19226, sle_31726, sle_37804, sle_42614, sle_53384, sle_53384 and sle_57074*) and 16S rRNA genes of *S. coelicolor* (*scor03, scor06, scor09, scor12, scor13 and scor16*). Blocks numbered in Roman indicate those regions where the amplified region perfectly matched with the sequence of the 16S rRNA genes of *S. coelicolor* and not with those of *S. leeuwenhoekii* C34. (**D**) Electropherogram of the sequencing result confirms that the amplified region correspond exclusively to the 16S rRNA of *S. coelicolor* M1650, since there is no traces of nucleotides that correspond to the 16S rRNA gene of *S. leeuwenhoekii* C34.

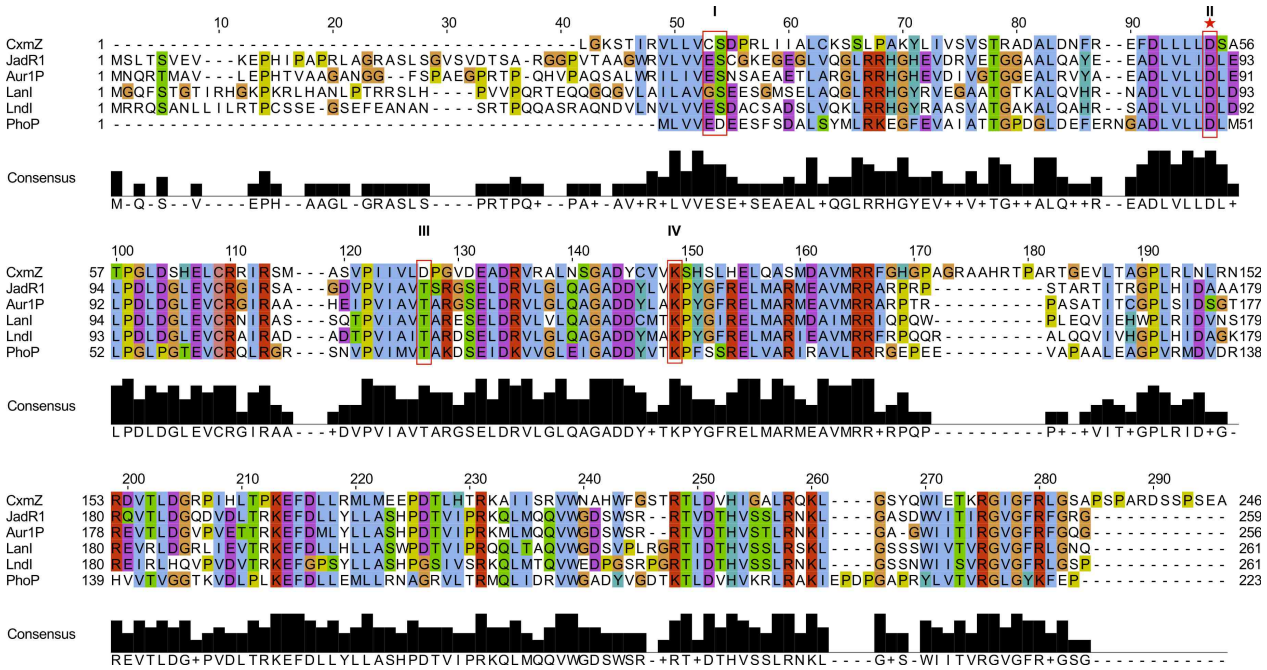


Figure H.10: Alignment of the amino acid sequence of putative transcriptional regulator CxmZ of *S. leeuwenhoekii* C34 with atypical RRs of the OmpR family of transcriptional activators. The amino acid sequences were obtained from the NCBI website, aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). Atypical RRs do not contain all the conserved amino acid residues for the phosphorylation of the receiver domain, such as aspartate in box I, and are likely activated by the presence of secondary metabolites (Liu et al., 2013). Red boxes indicate the residues present in typical RRs, like PhoP of *S. coelicolor* (CAB77324), which is phosphorylated and activated by its cognate sensor-kinase (Allenby et al., 2012). NCBI accession in parentheses. CxmZ of the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii* C34 (CQR60474); JadR1 of the jadomycin biosynthesis gene cluster of *S. venezuelae* (AAB36584); Aur1P of the auricin biosynthesis gene cluster of *S. aureofaciens* (CAI78073); LanI of the landomycin biosynthesis gene cluster of *S. cyanogenus* (AAO32359) and LndI of the landomycin E biosynthesis gene cluster of *S globisporus* (AAU04840).

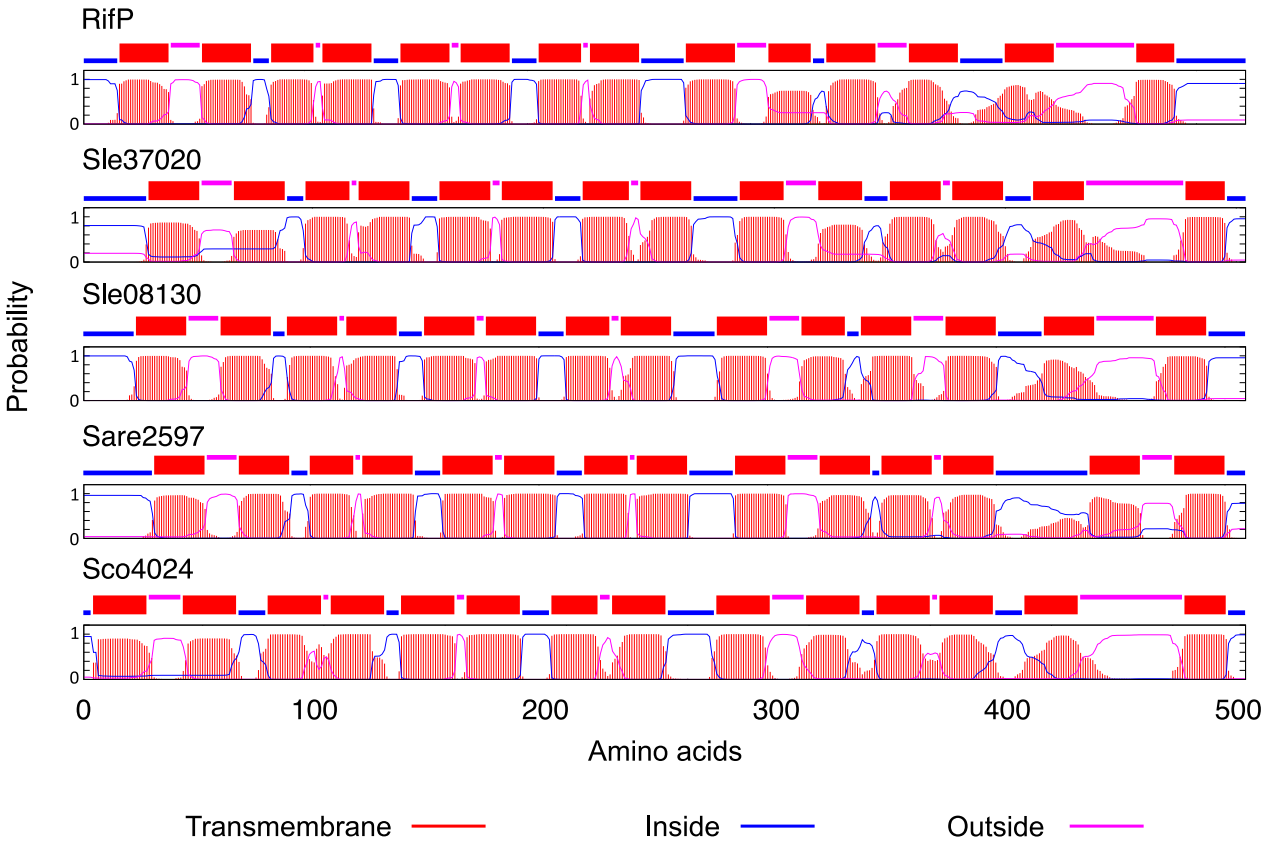


Figure H.11: Prediction of the transmembrane domains of the antibiotic transporter RifP of *Amycolatopsis mediterranei* and homologue transporters found in *Streptomyces leeuwenhoekii* C34. Transmembrane domains were predicted using TMHMM predictor server version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/; Krogh et al., 2001) and are indicated with red lines and blocks. NCBI accession in parentheses. RifP of *A. mediterranei* S699 (AAC01725) has been experimentally identified as one transporter of rifamycin B (Absalón et al., 2007). RifP shares high homology with Sle_37020 of *S. leeuwenhoekii* C34 (62%; antiseptic resistance protein; CQR63162); Sle_08130 of *S. leeuwenhoekii* C34 (60%; methyl viologen resistance protein smvA; CQR60276); Sare_2597 of *Salinispora arenicola* CNS-205 (58%; conserved hypothetical protein; ABV98438); Sco4024 of *S. coelicolor* (62%; integral membrane efflux protein; CAC32364). All of them are predicted to possess 14 transmembrane domains.

Appendix I

Supplementary information of Chapter Three: Identification of the putative locus of the chaxalactin biosynthesis gene cluster in the genome of *Streptomyces leeuwenhoekii* C34

I.1 Instructions for setting up RPS-BLAST analysis of proteins of *S. leeuwenhoekii* C34

Instructions for setting up RPS-BLAST analysis of proteins of *S. leeuwenhoekii* C34, searching for putative PKS domains. NCBI package 2.2.26+ (Camacho et al., 2009) is required for this analysis.

1. Content of the file "PKS_CDD.pn":

```
smart00822.smp
smart00823.smp
smart00824.smp
smart00825.smp
smart00826.smp
smart00827.smp
smart00829.smp
```

2. Create a RPS-type database in Command line:

```
makeprofiledb -dbtype rps -title cdd in pks -in PKS_CDD.pn -out
    PKS_CDD
```

   Where `makeprofiledb -dbtype rps` is for creating a RPS database for BLAST search analysis; `-title cdd in pks` is the name of the database; `-in PKS_CDD.pn` is the path to the input file with the list of the domains with which the database will be created; `-out PKS_CDD` is the path to the database.

3. Perform RPS-BLAST of all *S. leeuwenhoekii* C34 proteins:

```
rpsblast -query path/to/input -db PKS_CDD -out path/to/output.xml -
    outfmt 5 -evalue 0.00001
```

   Where `-query path/to/input` is the path to the input file with the protein sequences in fasta format; `-title cdd in pks` is the name of the database; `-db PKS_CDD` is the name of the database created above; `-out path/to/output.xml` is the path to the output file; `-outfmt 5` will format the output file in eXtensible Markup Language (xml), needed to parse the information in Python; `-evalue 0.00001` is the threshold of the expected-value.

4. Python code used to parse the .xml file with the output of the RPS-BLAST analysis (*next page*):

```
from Bio.Blast import NCBIXML #Import package for parsing
print ``ACP=SMART00823\t AT=SMART00827\t DH=SMART00826\t ER=SMART00829\t KR=
    SMART00822\t KS=SMART00825\t TE=SMART00824\n CDD\t SMART\t SHORT NAME\t E-
    value\t IDENTITIES\t FROM\t TO" #Header
for record in NCBIXML.parse(open("/Path/to/your/*.xml/file")) : #Begin Parse
    if record.alignments : #Do the following only if record.alignmet have a
        positive result in the *.xml file
        print ``\n"
        print ``Query: \%s..." \% record.query #[16:] #Print the name of the
            query sequence
        for align in record.alignments :
            for hsp in align.hsps :
                a = align.hit_def[:10] #Abbreviated hit name
                if a == str("smart00823"):
                    print `` \%s\t \%s\t ACP\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end) #print these fields from hps
                elif a == str("smart00827"):
                    print `` \%s\t \%s\t AT\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
                elif a == str("smart00826"):
                    print `` \%s\t \%s\t DH\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
                elif a == str("smart00829"):
                    print `` \%s\t \%s\t ER\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
                elif a == str("smart00822"):
                    print `` \%s\t \%s\t KR\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
                elif a == str("smart00824"):
                    print `` \%s\t \%s\t TE\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
                elif a == str("smart00825"):
                    print `` \%s\t \%s\t KS\t \%e\t \%d/\%s\t \%i\t \%i" \
                          \%(align.hit_id, align.hit_def[:10], hsp.expect, hsp.
                              identities,hsp.align_length,hsp.query_start, hsp.
                              query_end)
```

Table I.1: Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of *S. leeuwenhoekii* C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
Sle_61270	2-NPD_like (CD04730)	2-Nitropropane dioxygenase , one of the nitroalkane oxidizing enzyme families, catalyses oxidative denitrification of nitroalkanes to their corresponding carbonyl compounds and nitrites. NDP is a member of the NAD(P)H-dependent flavin oxidoreductase family that reduce a range of alternative electron acceptors. Most use FAD/FMN as a cofactor and NAD(P)H as electron donor. Some contain 4Fe-4S cluster to transfer electrons from FAD to FMN	13-250	$1.28e^{-88}$
	YrpB (COG2070)	NAD(P)H-dependent flavin oxidoreductase YrpB, nitropropane dioxygenase family [General function prediction only]	1-319	$1.82e^{-91}$
Sle_61280	Cyclase (PF04199)	Putative cyclase; proteins in this family are thought to be cyclase enzymes. They are found in proteins involved in antibiotic synthesis. However, they are also found in organisms that do not make antibiotics pointing to a wider role for these proteins. The proteins contain a conserved motif HxGTHxDxPxH that is likely to form part of the active site	59-253	$1.19e^{-19}$
Sle_61290	FC-FACS_FadD_like (CD05936)	Prokaryotic long-chain fatty acid CoA synthetases similar to *E. coli* FadD; this subfamily of the AMP-forming adenylation family contains *E. coli* FadD and similar prokaryotic fatty acid CoA synthetases. FadD was characterized as a long-chain fatty acid CoA synthetase. The gene fadD is regulated by the fatty acid regulatory protein FadR. Fatty acid CoA synthetase catalyses the formation of fatty acyl-CoA in a two-step reaction: the formation of a fatty acyl-AMP molecule as an intermediate, followed by the formation of a fatty acyl-CoA. This is a required step before free fatty acids can participate in most catabolic and anabolic reactions	31-544	0
	CaiC (COG0318)	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II [Lipid transport and metabolism, Secondary metabolites biosynthesis, transport and catabolism]	31-545	$3.17e^{-160}$
Sle_61300	FCS (CD05921)	Feruloyl-CoA synthetase (FCS); feruloyl-CoA synthetase is an essential enzyme in the feruloyl acid degradation pathway and enables some proteobacteria to grow on media containing feruloyl acid as the sole carbon source. It catalyses the transfer of CoA to the carboxyl group of ferulic acid, which then forms feruloyl-CoA in the presence of ATP and Mg^{2+}. The resulting feruloyl-CoA is further degraded to vanillin and acetyl-CoA. Feruloyl-CoA synthetase (FCS) is a subfamily of the adenylate-forming enzymes superfamily	5-171	$3.44e^{-03}$
	BACL_like (CD05929)	Bacterial Bile acid CoA ligases and similar proteins; bile acid-Coenzyme A ligase catalyses the formation of bile acid-CoA conjugates in a two-step reaction: the formation of a bile acid-AMP molecule as an intermediate, followed by the formation of a bile acid-CoA. This ligase requires a bile acid with a free carboxyl group, ATP, Mg^{2+}, and CoA for synthesis of the final bile acid-CoA conjugate. The bile acid-CoA ligation is believed to be the initial step in the bile acid 7alpha-dehydroxylation pathway in the intestinal bacterium *Eubacterium* sp.	155-506	$6.20e^{-172}$
	PRK08276 (CD05921)	Long-chain-fatty-acid−CoA ligase; validated	14-509	0
Sle_61310	TetR_N (PF00440)	Bacterial regulatory proteins, TetR family	18-63	$7.98e^{-08}$
	betaine_BetI (TIGR03384)	Transcriptional repressor BetI; BetI is a DNA-binding transcriptional repressor of the bet (betaine) regulon. In sequence, it is related to TetR (pfam00440). Choline, through BetI, induces the expression of the betaine biosynthesis genes betA and betB by derepression. The choline porter gene betT is also part of this regulon in *Escherichia coli*. Note that a different transcriptional regulator, ArcA, controls the expression of bet regulon genes in response to oxygen, as BetA is an oxygen-dependent enzyme. [Regulatory functions, DNA interactions]	15-198	$9.02e^{-06}$
Sle_61320	EamA (PF00892)	EamA-like transporter family; this family includes many hypothetical membrane proteins of unknown function. Many of the proteins contain two copies of the aligned region. The family used to be known as domain of unknown function 6	20-142	$1.38e^{-04}$
	EamA (PF00892)		174-299	$5.06e^{-09}$
	2A78	Carboxylate/Amino acid/Amine transporter; [Transport and binding proteins, Amino acids, peptides and amines]	22-294	$1.62e^{-22}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
Sle_61330	DDE_Tnp_1_2 (PF13586)	Transposase DDE domain; transposase proteins are necessary for efficient DNA transposition. This domain is a member of the DDE superfamily, which contain three carboxylate residues that are believed to be responsible for coordinating metal ions needed for catalysis	21-116	$6.77e^{-18}$
	DDE_Tnp_1 (PF01609)	Transposase DDE domain; transposase proteins are necessary for efficient DNA transposition. This domain is a member of the DDE superfamily, which contain three carboxylate residues that are believed to be responsible for coordinating metal ions needed for catalysis. The catalytic activity of this enzyme involves DNA cleavage at a specific site followed by a strand transfer reaction. This family contains transposases for IS*4*, IS*421*, IS*5377*, IS*427*, IS*402*, IS*1355*, IS*5*, which was original isolated in bacteriophage lambda	15-116	$3.93e^{-11}$
Sle_61340	DUF4096 (PF13340)	Putative transposase of IS*4/5* family (DUF4096)	10-83	$3.20e^{-31}$
	COG3293 (COG3293)	Transposase [Mobilome: prophages, transposons]	7-127	$1.87e^{-16}$
Sle_61350	TM_PBP2 (CD06261)	Transmembrane subunit (TM) found in Periplasmic Binding Protein (PBP)-dependent ATP-Binding Cassette (ABC) transporters which generally bind type 2 PBPs. These types of transporters consist of a PBP, two TMs, and two cytoplasmic ABC ATPase subunits, and are mainly involved in importing solutes from the environment. The solute is captured by the PBP which delivers it to a gated translocation pathway formed by the two TMs. The two ABCs bind and hydrolyze ATP and drive the transport reaction. For these transporters the ABCs and TMs are on independent polypeptide chains. These systems transport a diverse range of substrates. Most are specific for a single substrate or a group of related substrates; however some transporters are more promiscuous, transporting structurally diverse substrates such as the histidine/lysine and arginine transporter in *Enterobacteriaceae*. In the latter case, this is achieved through binding different PBPs with different specificities to the TMs. For other promiscuous transporters such as the multiple-sugar transporter MSM of *Streptococcus mutans*, the PBP has a wide substrate specificity. These transporters include the maltose-maltodextrin, phosphate and sulfate transporters, among others	94-249	$1.56e^{-13}$
	UgpE (COG0395)	ABC−type glycerol-3-phosphate transport system, permease component [Carbohydrate transport and metabolism]	21-304	$9.74e^{-69}$
Sle_61360	TM_PBP2 (CD06261)	See description above	95-292	$1.77e^{-08}$
	PotC (COG1177)	ABC−type spermidine/putrescine transport system, permease component II [Amino acid transport and metabolism]	20-308	$9.46e^{-11}$
Sle_61370	DDE_Tnp_1_6 super family (PF13586)	Transposase DDE domain; transposase proteins are necessary for efficient DNA transposition. This domain is a member of the DDE superfamily, which contain three carboxylate residues that are believed to be responsible for coordinating metal ions needed for catalysis	36-114	$5.74e^{-09}$
	DDE_Tnp_1 (PF01609)	Transposase DDE domain; Transposase proteins are necessary for efficient DNA transposition. This domain is a member of the DDE superfamily, which contain three carboxylate residues that are believed to be responsible for coordinating metal ions needed for catalysis. The catalytic activity of this enzyme involves DNA cleavage at a specific site followed by a strand transfer reaction. This family contains transposases for IS*4*, IS*421*, IS*5377*, IS*427*, IS*402*, IS*1355*, IS*5*, which was original isolated in bacteriophage lambda	1-113	$1.95e^{-09}$
Sle_61380	DUF4096 (PF13340)	Putative transposase of IS*4/5* family (DUF4096)	1-80	$1.47e^{-28}$
	COG3293 (COG3293)	Transposase [Mobilome: prophages, transposons]	1-125	$6.95e^{-13}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
Sle_61390	AdoMet_MTases (CD02440)	*S*-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use *S*-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product *S*-adenosyl-L-homocysteine (AdoHcy). There are at least five structurally distinct families of AdoMet-MTases, class I being the largest and most diverse. Within this class enzymes can be classified by different substrate specificities (small molecules, lipids, nucleic acids, etc.) and different target atoms for methylation (nitrogen, oxygen, carbon, sulphur, etc.)	55-158	$2.01e^{-14}$
	PKS_MT (SM00828)	Methyltransferase in PKS enzymes	53-265	$1.03e^{-16}$
CxlF (Sle_61400)	CypX (COG2124)	Cytochrome P450 [Secondary metabolites biosynthesis, transport and catabolism, Defense mechanisms]; linked to 3D-structure	8-374	$2.12e^{-37}$
CxlA (Sle_61410)	PKS (CD00833)	PKSs polymerize simple fatty acids into a large variety of different products, called polyketides, by successive decarboxylating Claisen condensations. PKSs can be divided into 2 groups, modular type I PKSs consisting of one or more large multifunctional proteins and iterative type II PKSs, complexes of several monofunctional subunits	21-426	0
	PKS_AT (SM00827)	Acyl transferase domain in PKS enzymes	531-825	$1.09e^{-122}$
	PKS_PP (SM00823)	Phosphopantetheine attachment site; phosphopantetheine (or pantetheine 4' phosphate) is the prosthetic group of acyl carrier proteins (ACP) in some multienzyme complexes where it serves as a 'swinging arm' for the attachment of activated fatty acid and amino acid groups	888-960	$1.46e^{-14}$
	PKS (CD00833)	See description above	987-1403	0
	PKS_AT (SM00827)	See description above	1500-1763	$2.69e^{-112}$
	PS-DH (PF14765)	Polyketide synthase dehydratase; this is the dehydratase domain of PKSs. Structural analysis shows these DH domains are double hotdogs in which the active site contains a histidine from the N-terminus hotdog and an aspartate from the C-terminus hotdog. Studies have uncovered that a substrate tunnel formed between the DH domains may be essential for loading substrates and unloading products	1842-2112	$6.45e^{-62}$
	KR_3_FAS_SDR_x (CD08956)	Beta-ketoacyl reductase (KR) domain of fatty acid synthase (FAS), subgroup 3, complex (x); ketoreductase, a module of the multidomain PKS, has 2 subdomains, each corresponding to a SDR family monomer. The C-terminus subdomain catalyses the NADPH-dependent reduction of the beta-carbonyl of a polyketide to a hydroxyl group, a step in the biosynthesis of polyketides, such as erythromycin. The N-terminus subdomain, an interdomain linker, is a truncated Rossmann fold which acts to stabilises the catalytic subdomain. Unlike typical SDRs, the isolated domain does not oligomerise but is composed of 2 subdomains, each resembling an SDR monomer. The active site resembles that of typical SDRs, except that the usual positions of the catalytic Asn and Tyr are swapped, so that the canonical YxxxK motif changes to YxxxN	2122-2548	$1.13e^{-145}$
	PKS_PP (SM00823)	See description above	2556-2640	$5.65e^{-28}$
	PKS (CD00833)	See description above	2657-3078	0
	PKS_AT (SM00827)	See description above	3175-3437	$4.61e^{-112}$
	KR_3_FAS_SDR_x (CD08956)	See description above	3551-4015	$3.56e^{-170}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
	PKS_PP (SM00823)	See description above	4031-4104	$1.30e^{-22}$
	PKS (CD00833)	See description above	4136-4559	0
	PKS_AT (SM00827)	See description above	4666-4929	$7.46e^{-112}$
	PS-DH (PF14765)	See description above	5010-5277	$1.10e^{-59}$
	KR_3_FAS_SDR_x (CD08956)	See description above	5285-5696	$4.96e^{-108}$
	PKS_PP (SM00823)	See description above	34-454	$1.65e^{-25}$
CxlB (Sle_61430)	PKS (CD00833)	See description above	2657-3078	0
	PKS_AT (SM00827)	See description above	558-838	$1.22e^{-111}$
	PS-DH (PF14765)	See description above	906-1176	$3.84e^{-56}$
	KR_3_FAS_SDR_x (CD08956)	See description above	1186-1629	$5.33e^{-132}$
	PKS_PP (SM00823)	See description above	1649-1734	$6.10e^{-26}$
	PKS (CD00833)	See description above	1753-2174	0
	PKS_AT (SM00827)	See description above	2270-2552	$2.15e^{-111}$
	KR_3_FAS_SDR_x (CD08956)	See description above	2656-3109	$5.29e^{-154}$
	PKS_PP (SM00823)	See description above	3129-3214	$5.54e^{-26}$
CxlC (Sle_61440)	PKS (CD00833)	See description above	35-456	0
	PKS_AT (SM00827)	See description above	563-859	$8.62e^{-123}$
	KR_3_FAS_SDR_x (CD08956)	See description above	961-1445	$1.03e^{-170}$
	PKS_PP (SM00823)	See description above	1463-1548	$8.27e^{-32}$
	PKS (CD00833)	See description above	1567-1988	0
	PKS_AT (SM00827)	See description above	2090-2371	$1.91e^{-113}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
	PS-DH (PF14765)	See description above	2428-2693	$2.79e^{-63}$
	KR_3_FAS_SDR_x (CD08956)	See description above	2699-3120	$1.37e^{-126}$
	PKS_PP (SM00823)	See description above	3142-3220	$4.89e^{-25}$
CxlD (Sle_61450)	PKS (CD00833)	See description above	39-465	0
	PKS_AT (SM00827)	See description above	573-867	$9.63e^{-111}$
	PS-DH (PF14765)	See description above	937-1205	$7.69e^{-60}$
	KR_3_FAS_SDR_x (related to enoyl reductase domain; CD08956)	In some instances, such as porcine FAS, an enoyl reductase (ER) module is inserted between the sub-domains	1208-1378	$9.00e^{-30}$
	enoyl_red (CD05195)	Enoyl reductase of PKS; putative enoyl reductase of PKS. PKSs produce polyketides in step by step mechanism that is similar to fatty acid synthesis. Enoyl reductase reduces a double to single bond. Erythromycin is one example of a polyketide generated by 3 complex enzymes (megasynthases)	1409-1690	$1.17e^{-83}$
	KR_3_FAS_SDR_x (CD08956)	See description above	1692-1951	$2.50e^{-87}$
	PKS_PP (SM00823)	See description above	1966-2051	$4.50e^{-31}$
	PKS (CD00833)	See description above	2071-2491	0
	PKS_AT (SM00827)	See description above	2588-2869	$1.13e^{-107}$
	PS-DH (PF14765)	See description above	2937-3210	$1.32e^{-58}$
	KR_3_FAS_SDR_x (CD08956)	See description above	3201-3608	$4.84e^{-79}$
	PKS_PP (SM00823)	See description above	3621-3706	$1.27e^{-28}$
CxlE (Sle_61560)	PKS (CD00833)	See description above	34-457	0
	PKS_AT (SM00827)	See description above	564-859	$1.50e^{-118}$
	PS-DH (PF14765)	See description above	925-1196	$8.25e^{-63}$
	KR_3_FAS_SDR_x (CD08956)	See description above	1198-1630	$7.40e^{-150}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	Domain (accession)	Description of the domain	Interval	E-value
	PKS_PP (SM00823)	See description above	1647-1731	$5.57e^{-28}$
	PKS (CD00833)	See description above	1748-2162	0
	PKS_AT (SM00827)	See description above	2256-2516	$1.95e^{-113}$
	PS-DH (PF14765)	See description above	2590-2853	$6.43e^{-48}$
	KR_3_FAS_SDR_x (CD08956)	See description above	2857-3280	$3.41e^{-116}$
	PKS_PP (SM00823)	See description above	3290-3375	$3.48e^{-27}$
	PKS_TE (SM00824)	Thioesterase; peptide synthetases are involved in the non-ribosomal synthesis of peptide antibiotics. Next to the operons encoding these enzymes, in almost all cases, are genes that encode proteins that have similarity to the type II fatty acid thioesterases of vertebrates. There are also modules within the peptide synthetases that also share this similarity. With respect to antibiotic production, thioesterases are required for the addition of the last amino acid to the peptide antibiotic, thereby forming a cyclic antibiotic. Thioesterases (non-integrated) have molecular masses of 25-29 kDa	3444-3648	2.60^{-59}
Sle_61470	DDE_5 super family (CL17874)	DDE superfamily endonuclease; this family of proteins are related to pfam00665 and are probably endonucleases of the DDE superfamily. Transposase proteins are necessary for efficient DNA transposition. This domain is a member of the DDE superfamily, which contain three carboxylate residues that are believed to be responsible for coordinating metal ions needed for catalysis. The catalytic activity of this enzyme involves DNA cleavage at a specific site followed by a strand transfer reaction	3-31	$3.24e^{-05}$
Sle_61480[1]				
Sle_61490	Rieske_RO_-Alpha_N (CD03469)	Rieske non-haem iron oxygenase (RO) family, N-terminus Rieske domain of the oxygenase alpha subunit; The RO family comprise a large class of aromatic ring-hydroxylating dioxygenases found predominantly in microorganisms. These enzymes enable microorganisms to tolerate and even exclusively utilize aromatic compounds for growth. ROs consist of two or three components: reductase, oxygenase, and ferredoxin (in some cases) components. The oxygenase component may contain alpha and beta subunits, with the beta subunit having a purely structural function. Some oxygenase components contain only an alpha subunit. The oxygenase alpha subunit has two domains, an N-terminus Rieske domain with an [2Fe-2S] cluster and a C-terminus catalytic domain with a mononuclear Fe(II) binding site. The Rieske [2Fe-2S] cluster accepts electrons from the reductase or ferredoxin component and transfers them to the mononuclear iron for catalysis. Reduced pyridine nucleotide is used as the initial source of two electrons for dioxygen activation	57-167	$8.23e^{-29}$
Sle_61500	Rieske_RO_-Alpha_N (CD03469)	See description above	57-167	$1.06e^{-28}$
Sle_61510	AdoMet_MTases (CD02440)	*S*-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use *S*-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product *S*-adenosyl-L-homocysteine (AdoHcy). There are at least five structurally distinct families of AdoMet-MTases, class I being the largest and most diverse. Within this class enzymes can be classified by different substrate specificities (small molecules, lipids, nucleic acids, etc.) and different target atoms for methylation (nitrogen, oxygen, carbon, sulphur, etc.)	70-174	$7.89e^{-19}$
	Methyltransf_31 (Pfam13847)	Methyltransferase domain; this family appears to be have methyltransferase activity	65-212	$1.77e^{-27}$

Continued. Summary of conserved domains found in proteins encoded by the chaxalactin biosynthesis gene cluster of S. leeuwenhoekii C34.

CDS	**Domain** (accession)	**Description of the domain**	**Interval**	**E-value**
Sle_61520[1]	-			
Sle_61530[1]	-			
Sle_61540[1]	-			
Sle_61550[1]	-			

This Table summarises the protein domains encoded in the putative chaxalactin biosynthesis gene cluster (*cxl*) of *S. leeuwenhoekii* C34. The domains were identified by using the CDD Search at NCBI (Marchler-Bauer et al., 2015) (accessed on 24 September 2015). PKS Modules present in CxlA, CxlB, CxlC, CxlD and CxlE are marked in grey and white.
[1] No domain found.

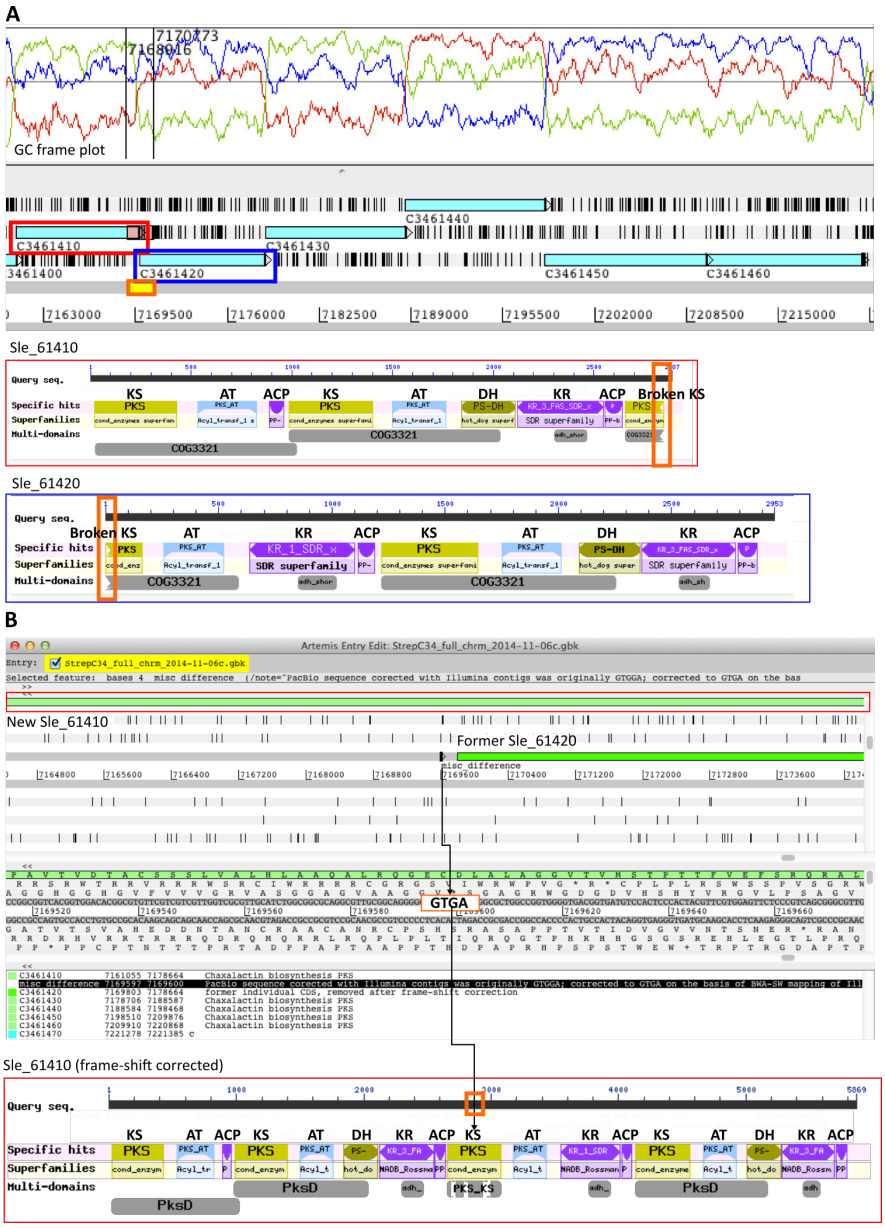


Figure I.2: Identification of a reading frame-shift in the putative chaxalactin PKS of *S. leeuwenhoekii* C34. (**A**) Two coding sequences, Sle_61410 and Sle_61420, were predicted as part of the chaxalactin PKS. The architecture of the protein domains translated from those coding sequences showed a broken KS domain at the C-terminus of Sle_61410 and at the N-terminus of Sle_61420, which is unusual. GC frame plot (Bibb et al., 1984) revealed a region where a possible frame-shift could be located (contained in an orange box). (**B**) Alignment of all Illumina reads on the genome sequence of *S. leeuwenhoekii* C34 revealed that the nucleotide sequence GTGGA had an extra guanine. This was carried out using software: Burrows-Wheeler Aligner, Smith-Waterman (BWA-SW; Li and Durbin, 2010) for the alignment and SAMtools (Li et al., 2009) for the single nucleotide polymorphisms identification (Juan Pablo Gomez-Escribano, personal communication). Thus, GTGGA sequence was corrected to GTGA (contained in an orange box), which fixed the reading frame-shift and the domain architecture of the PKS was as expected. In the current version of the genome, a new Sle_61410 (5,869 amino acids) includes the former Sle_61410 (2,887 amino acids) and Sle_61420 (2,953 amino acids).

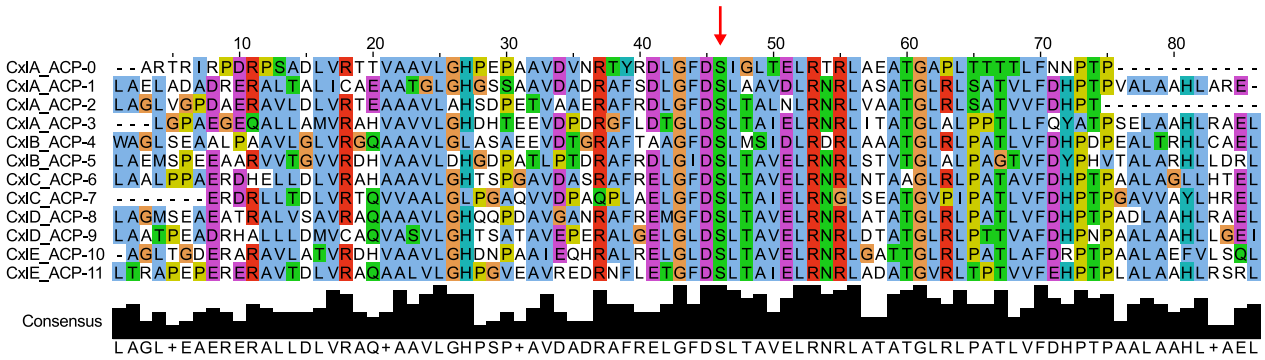


Figure I.3: Alignment of amino acid sequences of ACP domains of the chaxalactin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that ACP is located in the chaxalactin PKS ("CxlA_ACP-0", refers to the ACP of the Loading Module of the PKS subunit CxlA; "CxlA_ACP-1", refers to the ACP of Module 1 of the PKS subunit CxlA, and so on). The arrow indicates the position of the conserved serine residue that harbours the 4'-phosphopantetheine prosthetic group, which is a diagnostic residue to infer functionality of this domain.

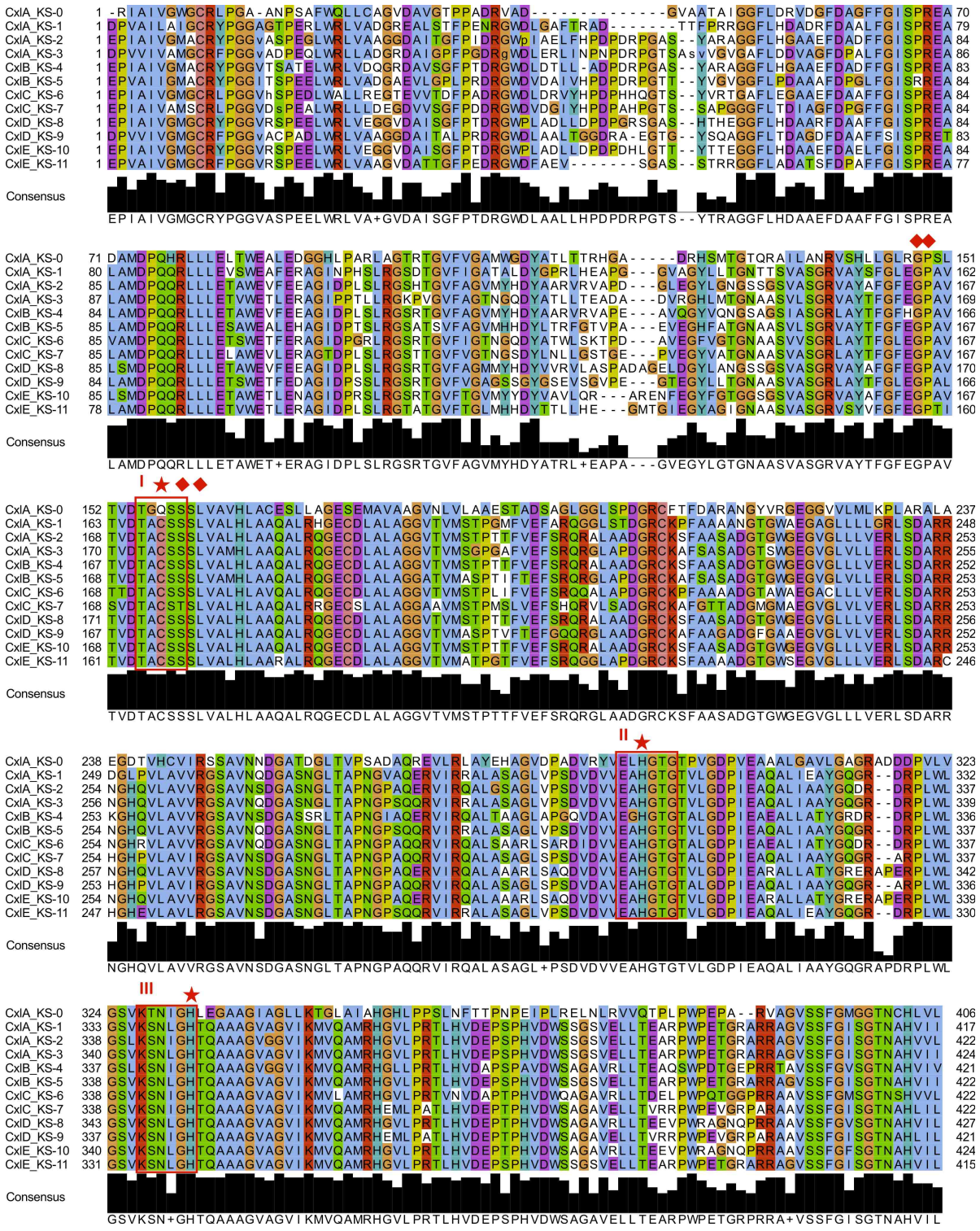


Figure I.4: Alignment of amino acid sequences of KS domains of the chaxalactin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that KS is located in the chaxalactin PKS ("CxlA_KS-0", refers to the KS of the Loading Module of the PKS subunit CxlA; "CxlA_KS-1", refers to the KS of Module 1 of the PKS subunit CxlA, and so on). The residues of the conserved motif GPxxxxxxC(S/A/T)(S/T)x(L/V) are indicated with a red diamond (Fernández-Moreno et al., 1992); catalytic triad cysteine, histidine and histidine present in motifs I, II and III, respectively are indicated with red stars (Keatinge-Clay, 2012).

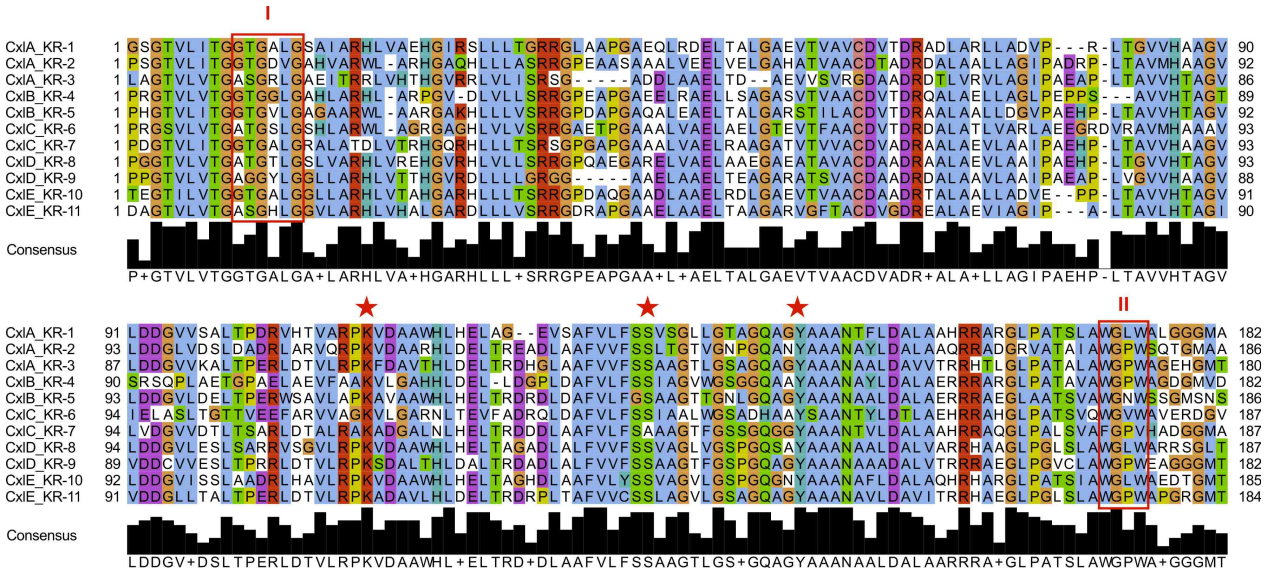


Figure I.5: Alignment of amino acid sequences of KR domains of the chaxalactin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that KR is located in the chaxalactin PKS ("CxlA_KR-1", refers to the KR of Module 1 of the PKS subunit CxlA; "CxlA_KR-2", refers to the KR of Module 2 of the PKS subunit CxlA, and so on). Reductase-competent KR domains contain the conserved motifs that allow NADPH binding, GxGxxG (I) and WGxW (II) which are contained in red boxes. The catalytic triad in the active site, lysine, serine and tyrosine are indicated with red stars as in Reid et al., 2003.

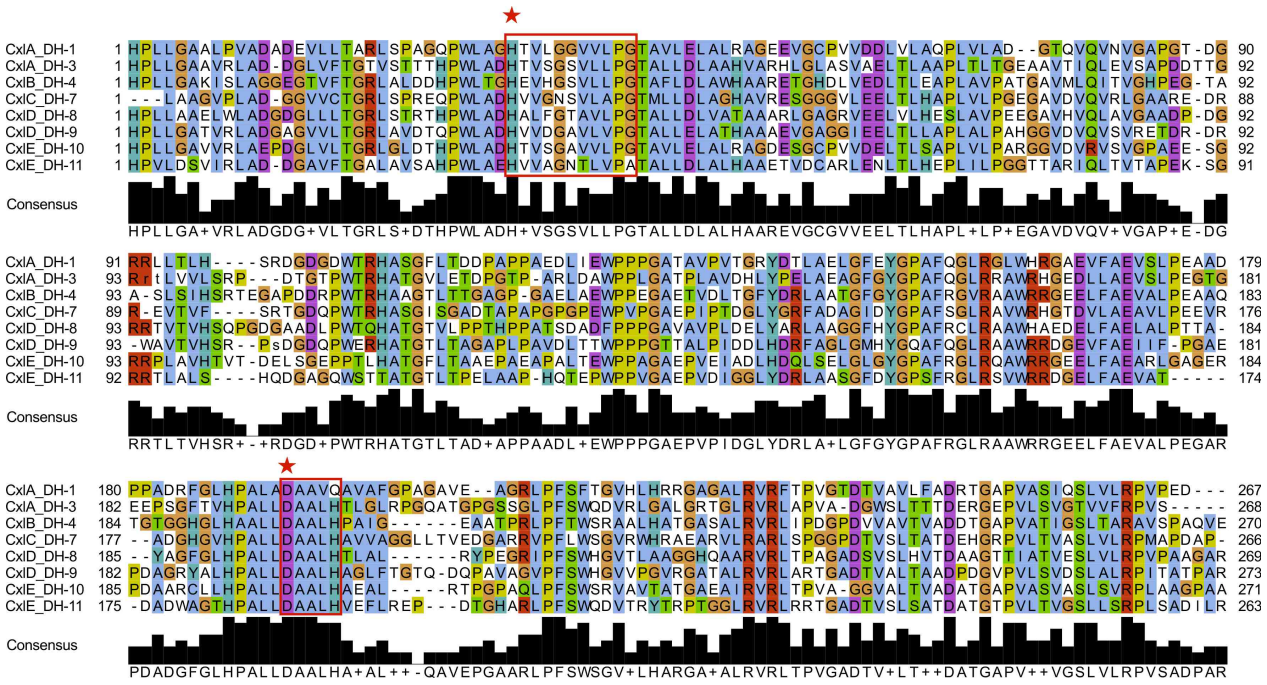


Figure I.6: Alignment of amino acid sequences of DH domains of the chaxalactin PKS. The amino acid sequences were identified by CDD Search at NCBI (Marchler-Bauer et al., 2015), aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009). The labels on the left indicate the module where that DH is located in the chaxalactin PKS ("CxlA_DH-1", refers to the DH of Module 1 of the PKS subunit CxlA; "CxlA_DH-3", refers to the DH of Module 3 of the PKS subunit CxlA, and so on). The conserved motifs are shown contained in red boxes and residues histidine and aspartate of the catalytic dyad are indicated with red stars.

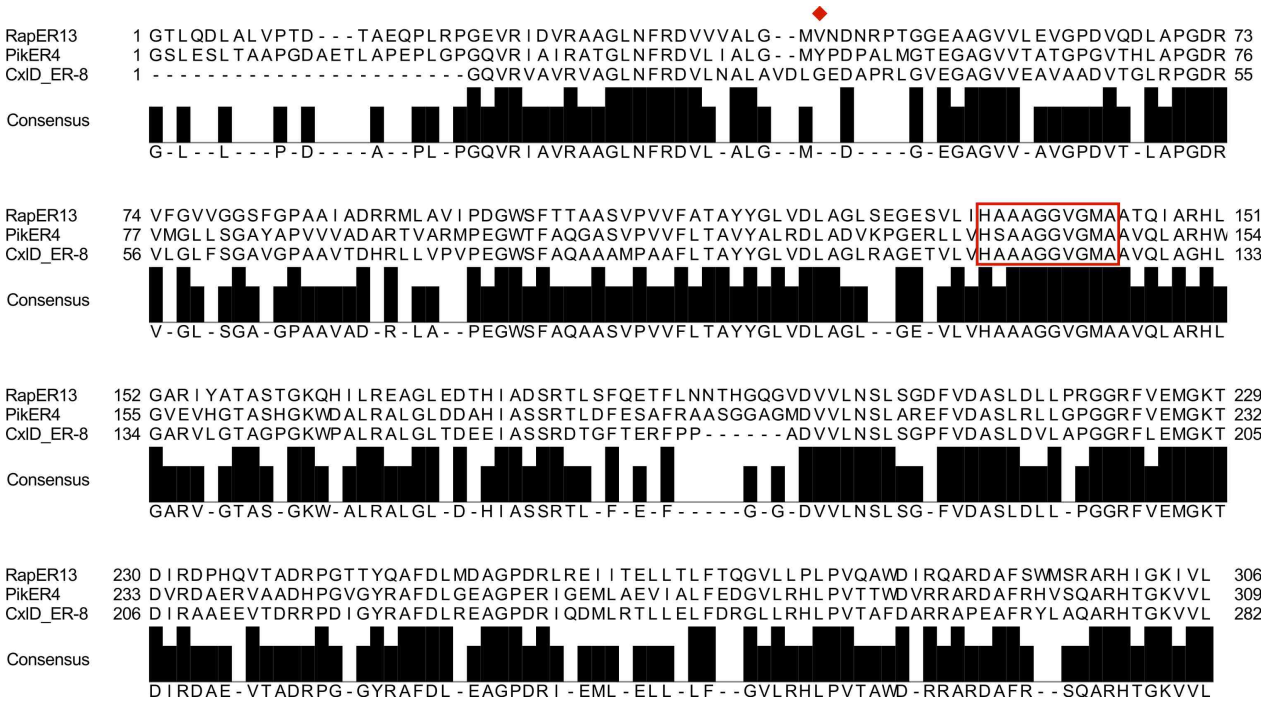


Figure I.7: Alignment of amino acid sequence of the ER domain of the chaxalactin PKS with those ER domains of known stereocontrol. The amino acid sequence in the chaxalactin ER domain of Module 8 of the PKS (CxlD_ER-8) was identified by CDD Search at NCBI (Marchler-Bauer et al., 2015). The sequences of ER of known stereocontrol were acquired from DoBISCUIT (Database of BIoSynthesis clusters CUrated and InTegrated; Ichikawa et al., 2013). ER domain of Module 4 of the pikromycin PKS (PikER4; accession: Pikro_00040) is known to leave the α-carbon in *S*-configuration (the α-substituent in L-orientation), which is correlated with the presence of a tyrosine residue (red diamond). ER domain of Module 13 of the rapamycin PKS (RapER13; accession: Rapam_00140) is known to leave the α-carbon in *R*-configuration (the α-substituent in D-orientation), which is correlated with the absence of a tyrosine residue (red diamond). The conserved NADPH-binding motif HxAx(G/T)GV(G/A)(M/S)A, is shown contained in a red box (Keatinge-Clay, 2012; Kwan et al., 2008). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

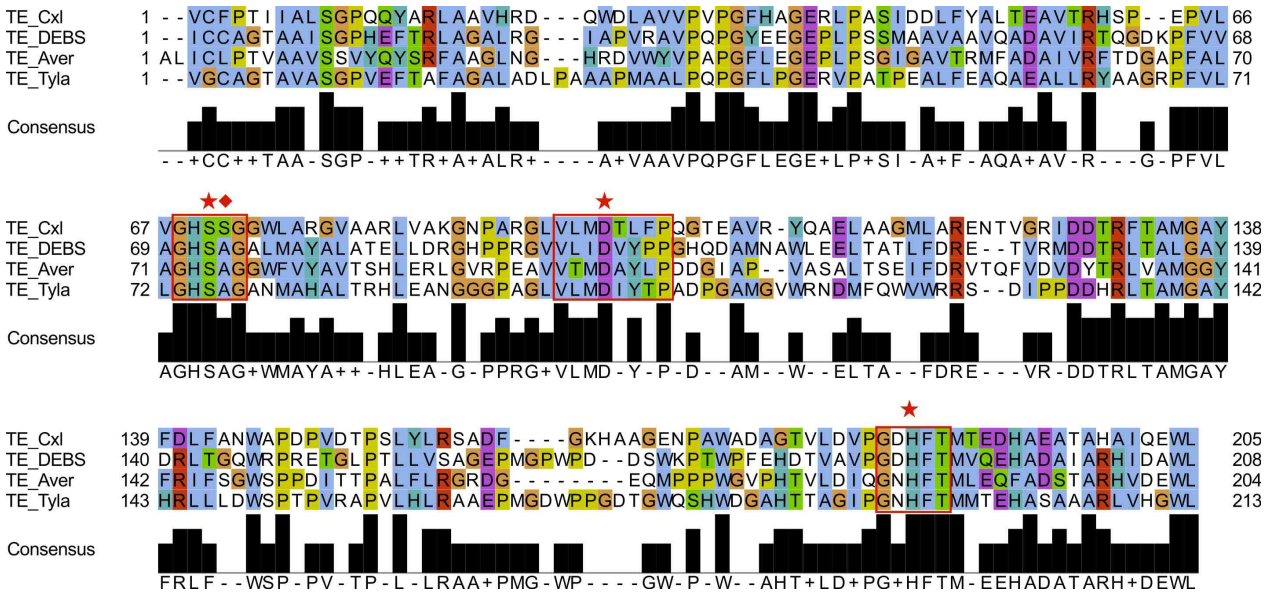


Figure I.8: Alignment of amino acid sequence of the TE domain of the chaxalactin PKS with similar TE domains of other type I PKSs. Residues serine, aspartate and histidine compose the catalytic triad, which are indicated with red stars and the stabilising residue is indicated with a red diamond (Tsai et al., 2001). Red boxes indicate conserved motifs around the active sites. The amino acid sequence of the TE domain of the chaxalactin PKS (Cxl_TE) was identified by CDD Search at NCBI (Marchler-Bauer et al., 2015). The sequences of TE of other type I PKSs were acquired from NCBI (accession): TE_DEBS = TE of DEBS of *Saccharopolyspora erythraea* NRRL 2338 (CAA44449); TE_Tyla = TE of tylactone synthase of *Streptomyces fradiae* (AAB66508); TE_Aver = TE of avermectin synthase of *Streptomyces avermitilis* (BAA84479). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

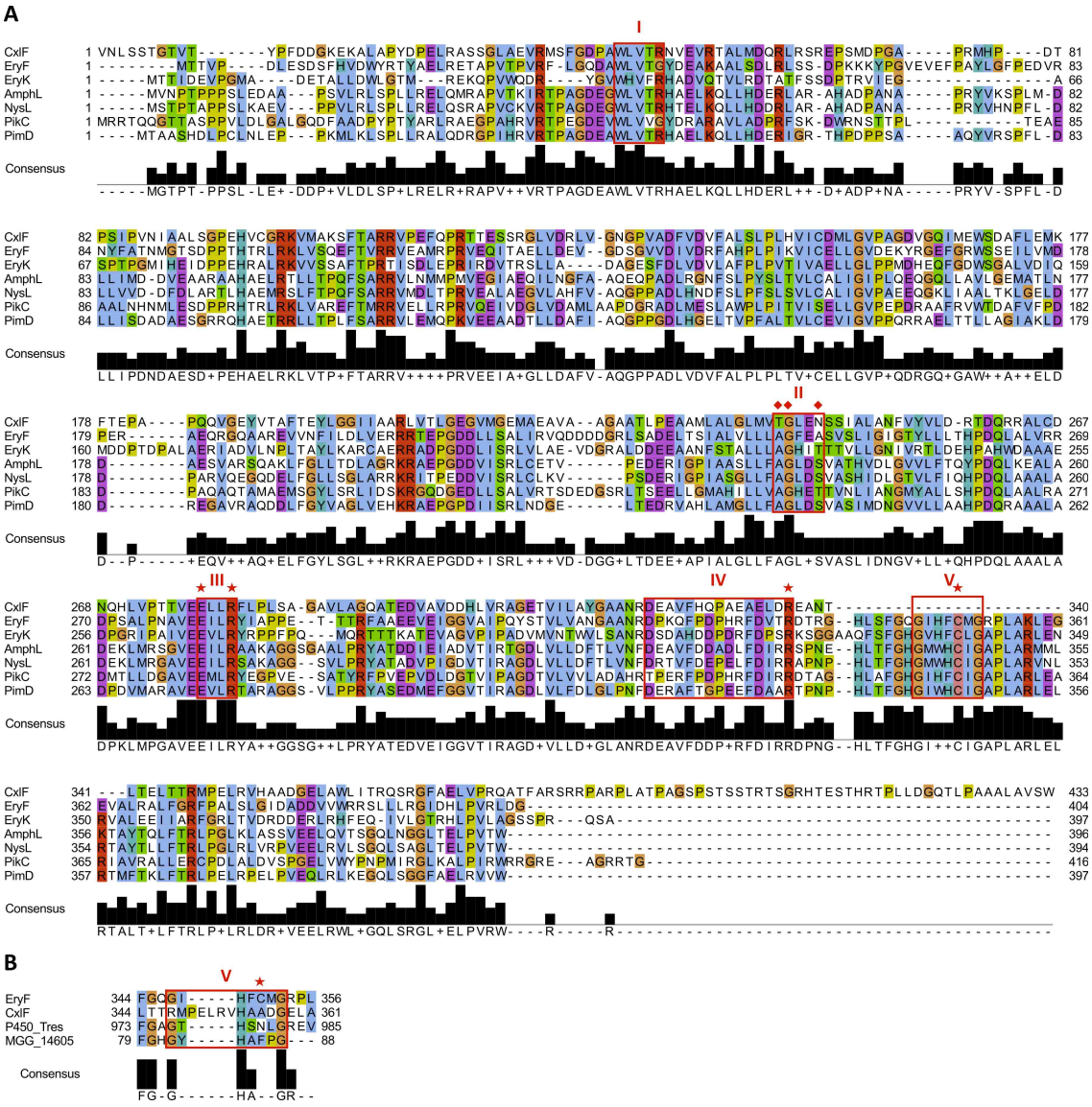


Figure I.9: Alignment of the amino acid sequences of the cytochrome P450 (CxlF) with other cytochromes P450. (**A**) Residues that make up the haem-binding pocket in cytochromes P450 are indicated with red stars and contained in the motifs (red boxes): WxxxR (I), **E**xx**R** (III), DxxFxxPxxFDxx**R** (IV) and Gxxx**C**xG (V). The cysteine in motif V is absent in CxlF. The oxygen binding and activation motif, A/G-GxxT (II), is present as TGxxN (residues indicated with red diamonds). (**B**) Alignment of a region of CxlF with cytochromes P450 devoid of the cysteine residue (Chen et al., 2014), and EryF that possesses the conserved cysteine residue (control). CxlF has an alanine instead of a cysteine in the corresponding position of the GxxxCxG motif. CxlF = Sle_61400 a cytochrome P450-SU1 of *Streptomyces leeuwenhoekii* C34 (CQR65596); EryF = 6-deoxyerythromolide B hydroxylase of *Saccharopolyspora erythrea* (AAA26496); EryK = erythromycin D C-12 hydroxylase of *Sacc. erythrea* (P48635); AmphL = amphotericin C-8 hydroxylase of *S. nodosus* (AJE39058); NysL = nystatin C-10 hydroxylase of *S. noursei* (AAF71769); PikC = narbomycin C-12 hydroxylase of *S. venezuelae* (O87605); PimD = 4,5-de-epoxypimaricin C-4-C-5 epoxidase of *S. natalensis* (CAC20932); P450_Tres is putative protein of *Trichoderma reesei* QM6a (EGR45174); MGG_14605 = hypothetical protein of *Magnaporthe oryzae* 70-15 (EHA56235). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

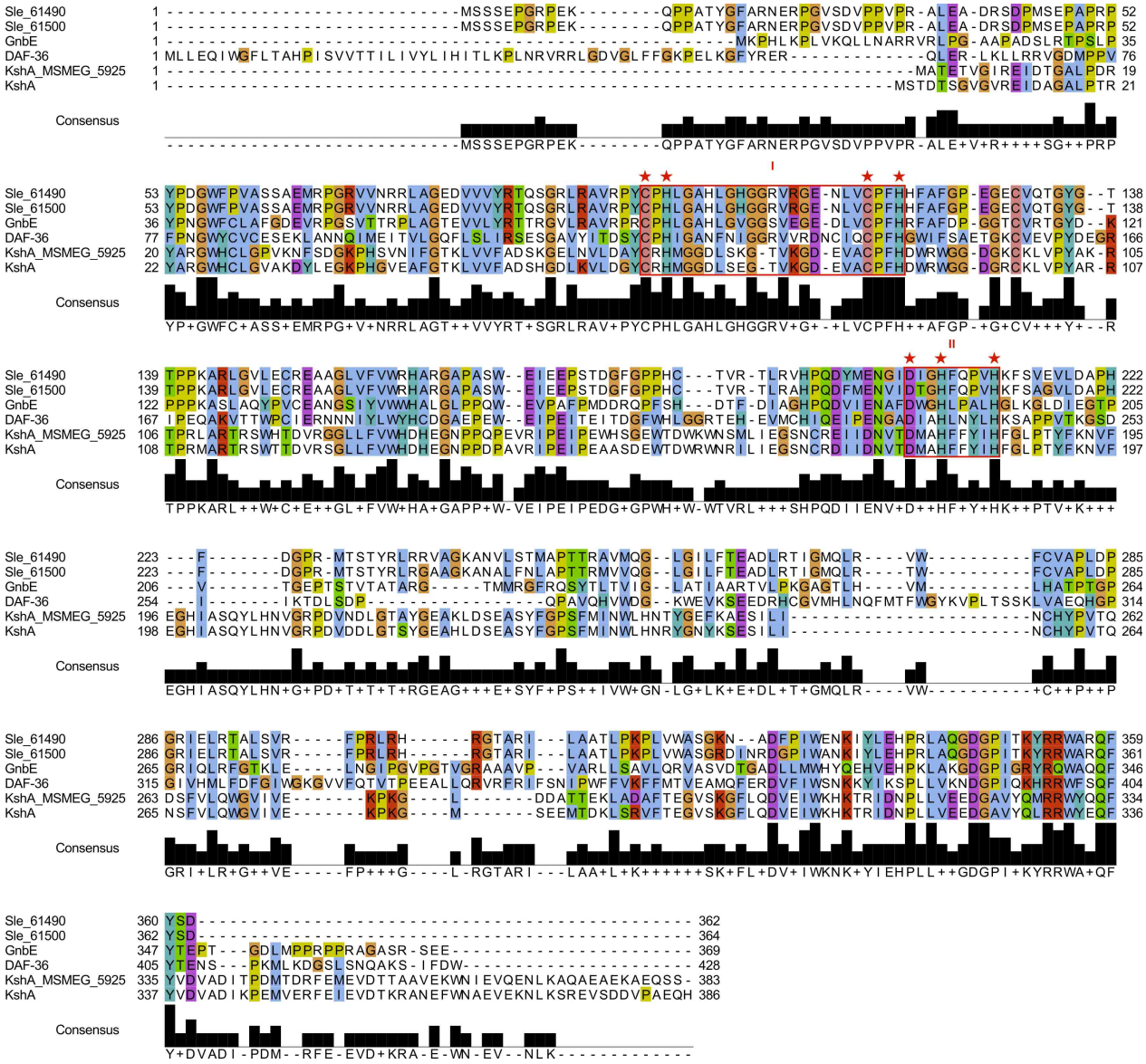


Figure I.10: Alignment of the amino acid sequences of Rieske domain-containing proteins Sle_61490 and Sle_61500 with characterised oxygenases. Conserved motifs present in Rieske domain-containing proteins (red boxes) include the **C**x**H**x_{17}**C**xx**H** that contains two cysteine and two histidine residues for binding the iron−sulphur, [2Fe−2S], cluster (I), and the **D**/E-x**H**xxxx**H** (II) that possesses an aspartate and two histidine residues for binding the non-haem iron atom (residues are indicated with red stars). Both motifs are needed for the oxygenase activity (Sydor et al., 2011). Sequences were obtained from NCBI (accession): Sle_61490 and Sle_61500 are Rieske-containing proteins of *Streptomyces leeuwenhoekii* C34 (CQR65604 and CQR65605, respectively); GnbE = Rieske iron-sulphur protein of *Streptomyces* sp. LZ35 (AJF48852); DAF-36 = cholesterol desaturase daf-36 of *Caenorhabditis elegans* (Q17938); KshA_MSMEG_5925 and KshA are the oxygenase subunits of the 3-ketosteroid-9α-monooxygenase of *Mycobacterium smegmatis* and *M. tuberculosis*, respectively (A0R4R3 and P71875, respectively). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

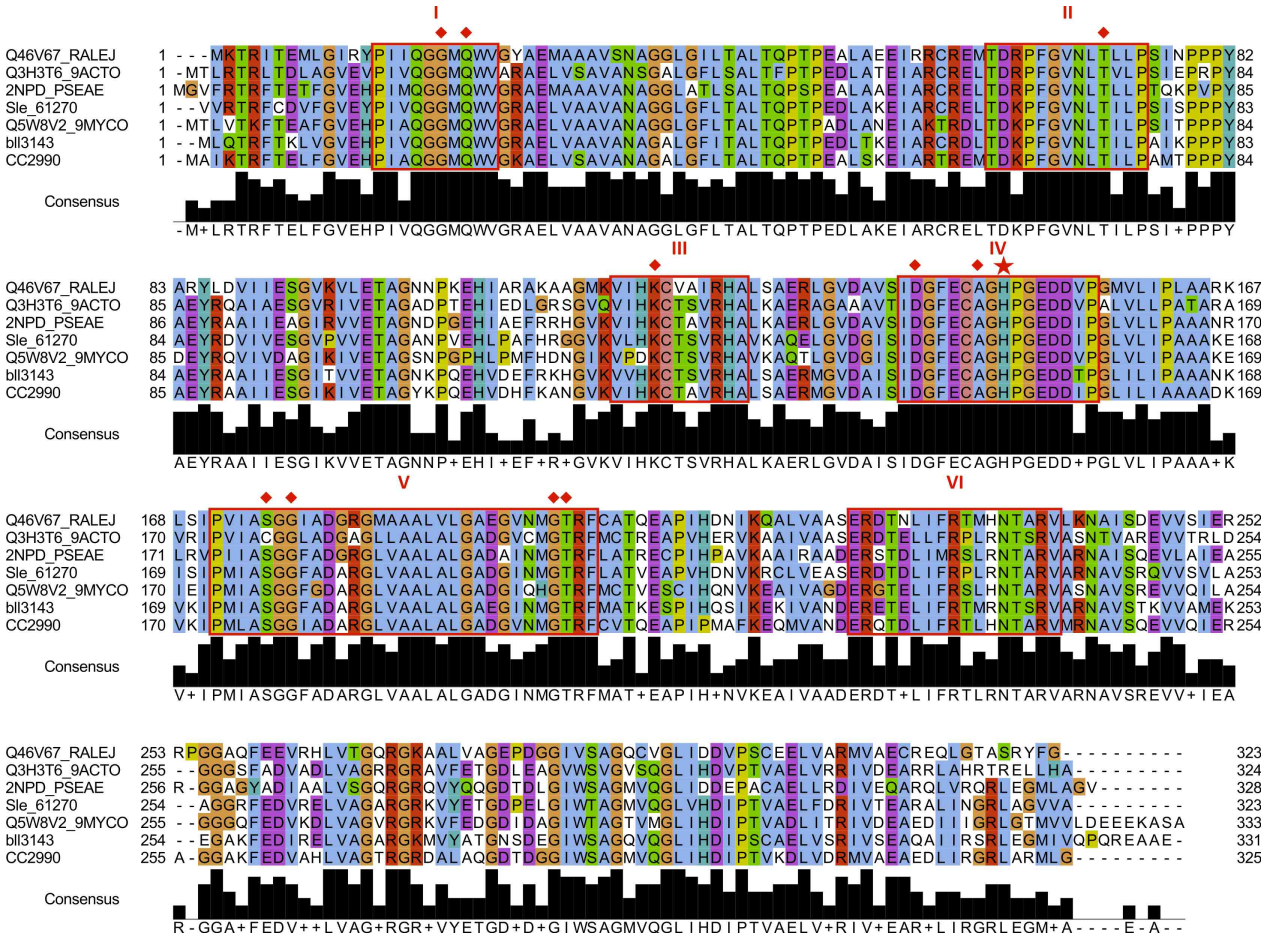


Figure I.11: Alignment of the amino acid sequence of Sle_61270, a putative 2-nitropropane dioxygenase, with homologues from other organisms. Residues that interact with FMN are indicated with red diamonds and the proposed catalytic base, histidine, is indicated with a red star. Motifs are contained in red boxes (Ha et al., 2006): **PI**h**QQGMQWV** (I), **TD**(K/R)**PFGVNLT**h**LP** (II), **V**hH**KC**TxV**RHA** (III), **IDGFECAGH-PGEDD**x**P** (IV), **ASGG**hA**D**(A/G)R**G**Lh**AA**L**ALG**A(D/E)GhxM**GTRF** (V), **ER**x**T**x**L**IF**R**x(L/M)(R/H)**NT**(S/A)**RV** (VI). Sle_61270 of *S. leeuwenhoekii* C34 (WP_029384442); 2NPD_PSEAE of *Pseudomonas aeruginosa* (Q9I4V0); bll3143 of *Bradyrhizobium diazoefficiens* (NP_769783); CC2990 of *Caulobacter crescentus* CB15 (Q9A453); Q3H3T6_9ACTO of *Nocardioides* sp. JS614 (Q3H3T6); Q46V67_RALEJ of *Ralstonia eutropha* strain JMP134 (Q46V67) and Q5W8V2_9MYCO of *Mycobacterium* sp. P101 (9MYCO). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

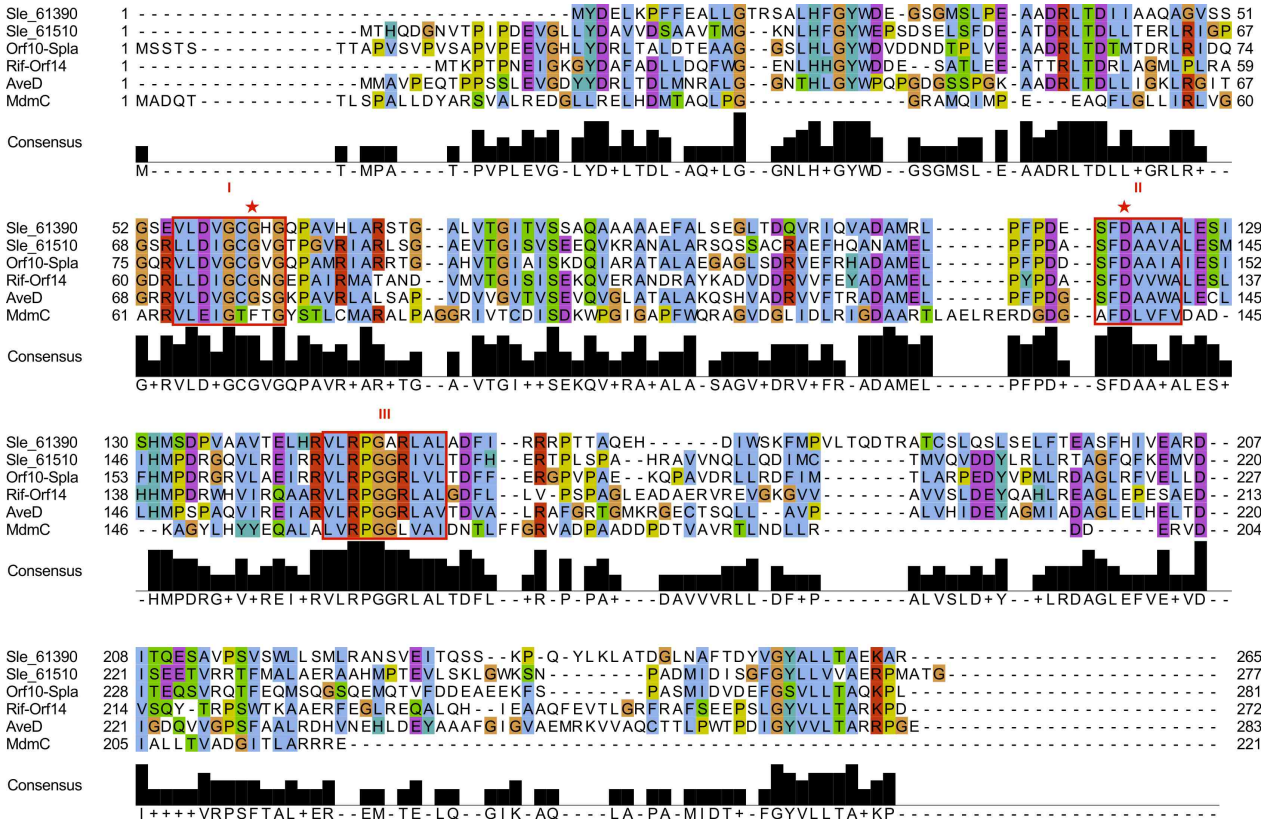


Figure I.12: Alignment of the amino acid sequences of Sle_61390 and Sle_61510, putative *O*-methyltransferases, with other methyltransferases of known substrate. Conserved motifs I, II and III were obtained from Kagan and Clarke, 1994 and are contained in red boxes. Motif I is (L/I/V)(V/L)(E/D)(V/I)**G**(C/G)**G**(P/T)**G** and contains the GxGxG sequence that is the nucleotide-binding site (SAM) (Schubert et al., 2003). Motif II is (G/P)(T/Q)(A/Y/F)**DA**(Y/V/I)(I/F)(L/V/C) in which the central aspartate residue is invariant, since it forms hydrogen bonds to both hydroxyls of the SAM ribose (Schubert et al., 2003). Motif III is LL(K/R)P**GG**(L/I/R)(I/L)(V/I/F/L)(L/I) whose central glycines are highly conserved and at least one of them in present across all methyltransferases. Amino acids indicated with red stars correspond to those highly conserved in PKS *O*-methyltransferases according to Ansari et al., 2008. Amino acid sequences were obtained from NCBI (accession): Sle_61390 and Sle_61510 = putative chaxalactin A C-13 *O*-methyltransferase of *Streptomyces leeuwenhoekii* C34 (CQR65595 and CQR65606, respectively); Orf10-Spla = putative *O*-methyltransferase of *Streptomyces platensis* subsp. *rosaceus* (ACY01395); Rif-Orf14 = C-27 *O*-demethylrifamycin SV methyltransferase of *Amycolatopsis mediterranei* S699 (AAC01738; Xu et al., 2003); MdmC = C-4 *O*-methyltransferase of *Streptomyces mycarofaciens* (AAA26782; Hara and Hutchinson, 1992); AveD = C-5 *O*-methyltransferase in avermectin biosynthesis of *Streptomyces avermitilis* (BAA84602; Ikeda et al., 1999). Sequences were aligned in ClustalX (Thompson et al., 1997) and visualised in Jalview (Waterhouse et al., 2009).

Appendix J

Research article *in extenso*: "Identification and heterologous expression of the chaxamycin biosynthesis gene cluster of *Streptomyces leeuwenhoekii*"

J. F. Castro, V. Razmilic, J. P. Gomez-Escribano, B. Andrews, J. A. Asenjo and M. J. Bibb.
Applied and Environmental Microbiology, 2015, volume 81, issue 17, pages 5820−5831.
DOI 10.1128/AEM.01039-15
Supplementary information available at AEM website:
http://aem.asm.org/content/81/17/5820/suppl/DCSupplemental.

AEM
Journals.ASM.org

Identification and Heterologous Expression of the Chaxamycin Biosynthesis Gene Cluster from *Streptomyces leeuwenhoekii*

Jean Franco Castro,[a,b] Valeria Razmilic,[a,b] Juan Pablo Gomez-Escribano,[b] Barbara Andrews,[a] Juan A. Asenjo,[a] Mervyn J. Bibb[b]

Centre for Biotechnology and Bioengineering, Department of Chemical Engineering and Biotechnology, Universidad de Chile, Santiago, Chile[a]; Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, United Kingdom[b]

***Streptomyces leeuwenhoekii*, isolated from the hyperarid Atacama Desert, produces the new ansamycin-like compounds chaxamycins A to D, which possess potent antibacterial activity and moderate antiproliferative activity. We report the development of genetic tools to manipulate *S. leeuwenhoekii* and the identification and partial characterization of the 80.2-kb chaxamycin biosynthesis gene cluster, which was achieved by both mutational analysis in the natural producer and heterologous expression in *Streptomyces coelicolor* A3(2) strain M1152. Restoration of chaxamycin production in a nonproducing Δ*cxmK* mutant (*cxmK* encodes 3-amino-5-hydroxybenzoic acid [AHBA] synthase) was achieved by supplementing the growth medium with AHBA, suggesting that mutasynthesis may be a viable approach for the generation of novel chaxamycin derivatives.**

Antibiotic resistance is rapidly becoming a worldwide medical problem that requires urgent attention (1–3). Historically, natural products produced by actinomycetes and fungi have been the major source of clinically used antibiotics (4). However, the repeated rediscovery of known chemical entities from natural sources has led to a marked decline in the discovery of novel microbial antibiotics with clinical utility (5, 6). One potential solution to this problem is to screen microorganisms isolated from previously little-scrutinized ecological niches, and bioprospecting of unexplored environments has indeed already resulted in the discovery of novel microorganisms and new chemical scaffolds (4, 6).

In 2004, a considerable number of new actinomycetes were, surprisingly, isolated from the hyperarid Atacama Desert, the oldest and driest desert on Earth, located in northern Chile (7, 8). Strikingly, some of these strains produce novel chemical structures with potent biological activity (9–11). One of the strains isolated from this environment is *Streptomyces leeuwenhoekii* DSM 42122 (previously known as *Streptomyces* strain C34^T), isolated from the Chaxa Lagoon (7, 12). It produces the new polyketide antibiotics called the chaxamycins (9) and the chaxalactins (10). Chaxamycin A to D (compounds 1 to 4 in Fig. 1) are four new naphthalenic ansamycin-type polyketides derived from a 3-amino-5-hydroxybenzoic acid (AHBA) starter unit. Chaxamycin D (compound 4) displays, through an unknown mechanism, highly selective antimicrobial activity against *Staphylococcus aureus* ATCC 25923 and against a panel of methicillin-resistant *S. aureus* (MRSA) strains (9), while chaxamycin A (compound 1) inhibits the intrinsic ATPase activity of the human Hsp90 protein, which is involved in cancer proliferation (9). The chaxamycins differ from other naphthalenic ansamycins, such as the rifamycins, naphthomycin, geldanamycin, and the herbimycins, in lacking a methyl group at the olefinic C-16 next to the amide bond present in all of these molecules.

The biosynthesis of ansamycin-type polyketides commences with priming of the polyketide synthase (PKS) with the precursor molecule AHBA. In the case of the rifamycin and other ansamycin-type polyketide gene clusters, a sub-gene cluster, located near the PKS, is largely responsible for the synthesis of AHBA (13). One of the enzymes encoded in the core of this AHBA gene cluster is AHBA synthase, which converts 5-deoxy-5-amino-3-dehydroshikimic acid (amino-DHS) into AHBA, the final step in its biosynthesis (14), and an AHBA synthase gene indeed was identified previously by PCR analysis of genomic DNA of *S. leeuwenhoekii* (9). In addition to AHBA, we predict that chaxamycin is synthesized from 10 additional extender units (three malonyl coenzyme As [malonyl-CoAs] and seven methylmalonyl-CoAs) by a modular type I PKS.

Recent improvements in next-generation sequencing technologies and genome mining have markedly accelerated the identification and characterization of gene clusters for specialized metabolites, many potentially encoding novel compounds with antimicrobial activity. In this work, we describe the application of these technologies to identify and partially characterize the chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii*.

MATERIALS AND METHODS

Strains, culture conditions, and general methods. The bacterial strains used and generated during this study are listed in Table 1. *Escherichia coli* and *Streptomyces* strains were cultured, maintained, and manipulated genetically following the methods of Sambrook et al. (15) and Kieser et al. (16), respectively. The plasmids and oligonucleotides used or constructed during this work are listed in Table 2 and Table 3, respectively. Mobilization of phage P1-derived artificial chromosome (PAC) clones into *Streptomyces* strains was performed as described previously (17). Molecular biology enzymes, reagents, and kits were used according to the manufac-

Received 1 April 2015 **Accepted** 12 June 2015

Accepted manuscript posted online 19 June 2015

Citation Castro JF, Razmilic V, Gomez-Escribano JP, Andrews B, Asenjo JA, Bibb MJ. 2015. Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from *Streptomyces leeuwenhoekii*. Appl Environ Microbiol 81:5820–5831. doi:10.1128/AEM.01039-15.

Editor: M. A. Elliot

Address correspondence to Mervyn J. Bibb, mervyn.bibb@jic.ac.uk.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01039-15.

doi:10.1128/AEM.01039-15

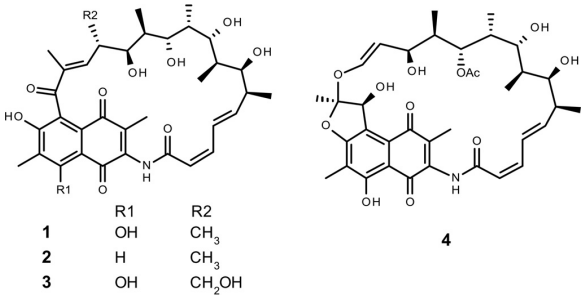


FIG 1 Structures of chaxamycins A to D (compounds 1 to 4).

turer's instructions. High-fidelity PCR amplification was performed with Phusion or Q5 DNA polymerase following the manufacturer's instructions (NEB, Ipswich, MA) with a nucleotide proportion of 15A:15T:35G:35C to improve the amplification efficiency of high-moles-percent G+C *Streptomyces* DNA. 3-Amino-5-hydroxybenzoic acid (AHBA) was purchased from Sigma (catalog no. PH011754) and dissolved in dimethyl sulfoxide.

Specific methods for *S. leeuwenhoekii*. Spore stocks of *S. leeuwenhoekii* were prepared following standard methods (16) from cultures grown on soy flour-mannitol (SFM) agar medium at 37°C for 5 to 7 days. Mobilization of plasmids into *S. leeuwenhoekii* was performed as established for *Streptomyces coelicolor* A3(2) by conjugation from *E. coli* ET12567 (carrying either pUZ8002 or pR9406) using 10^8 spores that were heat shocked in 2× YT medium at 50°C for 10 min; the treated spores were allowed to cool to room temperature before mixing them with the *E. coli* cells, and the conjugation mixtures were plated on SFM agar medium supplemented with 10 mM $MgCl_2$ and 10 mM $CaCl_2$.

Production and measurement of chaxamycins. *S. leeuwenhoekii* strains were grown in modified ISP2 medium (containing, per liter, 4.0 g of yeast extract, 10.0 g of malt extract, and 10.0 g of glycerol [pH 7.20] [9]). For heterologous production of chaxamycin, *S. coelicolor* strains were cultured in R3 medium (18) [containing, per liter, 10 g of glucose, 5 g of yeast extract, 100 mg of Casamino Acids, 3 g of proline, 10 g of $MgCl_2 \cdot 6H_2O$, 4 g of $CaCl_2 \cdot 2H_2O$, 200 mg of K_2SO_4, 50 mg of KH_2PO_4, 5.6 g of *N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES), and trace elements (as described for R2 medium in reference 16) and adjusted to pH 7.2 with NaOH]. All strains were cultured in 250-ml Erlenmeyer flasks that had been treated with 5% dimethyldichlorosilane (DMDCS) in toluene (Sylon CT; Supelco, catalog number 33065-U) to deter the mycelium from sticking to the glass; a stainless steel spring was placed at the bottom of each flask to improve aeration. Seed cultures were inoculated with 10^8 spores and incubated at 30°C with orbital shaking (250 rpm) for 24 to 48 h. Mycelium from the seed cultures was harvested by centrifugation, washed with 0.9% NaCl, resuspended in fresh medium, and used to inoculate production cultures to a final optical density at 600 nm (OD_{600}) of 0.2; production cultures were incubated at 30°C with orbital shaking (250 rpm). Samples were harvested at the indicated times: 1 or 2 ml of homogenized culture was pipetted into a clean Eppendorf tube and centrifuged at 13,000 rpm at 4°C for 10 min, the supernatant was moved to a new Eppendorf tube and centrifuged again at 13,000 rpm at 4°C for 10 min before carefully pipetting 0.5 ml of the supernatant into a high-pressure liquid chromatography (HPLC) vial, the mycelial pellet was extracted with 0.5 ml of methanol and centrifuged at 13,000 rpm at 4°C for 10 min, and 0.3 ml of the methanol extract was pipetted into an HPLC vial containing 0.3 ml of water. Chaxamycins were detected in both the supernatants and the mycelial extracts, but larger amounts were always found in the supernatants, and thus only these fractions were analyzed in most experiments. When analyzing total chaxamycin production in a single HPLC run, both fractions were mixed in a 1:1 ratio (thus injecting samples in 50% methanol). Samples were analyzed on a Shimadzu liquid chromatography-mass spectrometry (LC-MS) system equipped with a NexeraX2 liquid chromatograph (LC30AD) fitted with a Prominence photodiode array detector (SPD-M20A) and an LCMS-IT-ToF mass spectrometer; samples (typically 5 μl) were injected in a Kinetex XB C_{18} 2.6-μm, 100-Å, 50- by 2.10-mm column (part no. 00B-4496-AN; Phenomenex, USA) fitted with a KrudKatcher Ultra HPLC in-line filter (part no. AF0-8497; Phenomenex, USA) and eluted at a flow rate of 0.6 ml per min with a gradient of 0.1% formic acid in water (mobile phase A) and methanol (mobile phase B) as follows (minute, percentage of B reached at that minute): 0 min, 2% B; 1 min, 2% B; 8 min, 100% B; 9.3 min, 100% B; 9.5 min, 2% B; and 11.2 min, 2% B for equilibration. The column was kept at 40°C. Mass spectrometry detection was performed in negative ion mode since all of the chaxamycin species were detected with higher sensitivity than in positive mode. Purified chaxamycin standards for compounds 1, 2, and 3 were kindly provided by Mostafa Rateb and Marcel Jaspars (University of Aberdeen, Scotland). Data acquisition and analysis were performed with LCMSsolutions version 3 (Shimadzu); spectrum visualization was also performed with Mass++ version 2.7.2 (http://www.first-ms3d.jp/english/). The $(M - H)^-$ ions detected for the chaxamy-

TABLE 1 Bacterial strains used in this study

Species and strain	Description	Reference or source
Escherichia coli		
DH5α	Strain used for routine cloning	65
DH10B	Strain used for routine cloning	65
ET12567/pUZ8002	Methylation-deficient strain used for conjugation with *Streptomyces*; pUZ8002 provides conjugation machinery	66 (pUZ8002, J. Wilson and D. Figurski, unpublished)
TOP10/pR9406	Strain used for routine cloning carrying conjugation plasmid pR9406	pR9406, A. Siddique and D. Figurski, unpublished.
Streptomyces coelicolor		
M1152	M145 Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB*(C1298T)	67
M1650	M1152 carrying pIJ12853	This work
Streptomyces leeuwenhoekii		
Wild-type strain		12
M1653	*S. leeuwenhoekii* Δ*cxmK*::*neo*	This work
M1655	*S. leeuwenhoekii*/pGUS	This work
M1656	*S. leeuwenhoekii*/pIJ10740	This work

TABLE 2 Plasmids used and constructed during this study

Plasmid	Description	Reference or source
pBluescript II KS(+)	General cloning vector	68
pKC1132	Cloning vector, conjugative (*oriT* from RK2)	69
pSET152	Cloning vector, conjugative (*oriT* from RK2), integrative (phiC31 *attP*)	69
pTC192-Km	Source of *neo* (kanamycin resistance gene)	70
pGM1190	pSG5 derivative, *tsr*, *aac(3)IV*, *oriT*, *to* terminator, P_{tipA}, RBS, *fd* terminator	40 (pGM1190, G. Muth, unpublished)
pESAC13	PAC vector (P1 phage replicon) for genomic library construction; conjugative (*oriT* from RK2), integrative (phiC31 *attP*), *tsr*, *neo*, P1 *rep*, *sacB*	27 (pESAC13, M. Sosio, unpublished)
pGUS	*gusA* (β-glucuronidase gene, codon-optimized for streptomycetes) in pSET152	38
pIJ10740	pGUS derivative with *ermE**p driving *gusA* transcription	Morgan Feeney, unpublished
pIJ10257	Expression vector for *Streptomyces*, with *ermE**p, *hyg*, conjugative (*oriT* from RK2), integrative (phiC31 *attP*)	71
pIJ12850	Derivative of pGM1190 with JFC010/JFC009-*neo*-JFC011/JFC012 fragments; for deletion of *cxmK*	This work
pIJ12851	Derivative of pKC1132 with JFC010/JFC009-*neo*-JFC011/JFC012 fragments; for deletion of *cxmK*	This work
pIJ12857	Derivative of pKC1132 with JFC010/JFC009 fragment; for deletion of *cxmK*	This work
pIJ12858	Derivative of pKC1132 with JFC011/JFC012 fragment; for deletion of *cxmK*	This work
pIJ12859	Derivative of pKC1132 with JFC010/JFC009-JFC011/JFC012 fragment; for deletion of *cxmK*	This work
pIJ12860	Derivative of pGM1190 with JFC010/JFC009-JFC011/JFC012 fragment; for deletion of *cxmK*	This work
pIJ12861	pBluescript II KS(+) derivative with JFC026/JFC034 fragment; for cloning of *cxmK*	This work
pIJ12852	pIJ10257 derivative with *cxmK*; for complementation of deletion mutant	This work
pIJ12853	Derivative of pESAC13; contains 145 kb of *S. leeuwenhoekii* chromosome, including the chaxamycin biosynthesis gene cluster	This work

cin species were as follows: compound 1, *m/z* 638.29; compound 2, *m/z* 622.29; compound 3, *m/z* 654.29; and compound 4, *m/z* 682.29.

DNA sequencing and bioinformatic analysis. The sequencing and assembly of the genome sequence of *S. leeuwenhoekii* has been reported elsewhere (19). Confirmatory Sanger sequencing was performed using the Eurofins Genomics service (Ebersberg, Germany) and analyzed using the Staden package (20) version 2.0.0b9 (http://staden.sourceforge.net). BLAST (21) searches were performed at the NCBI server. Local BLAST searching was performed using NCBI package 2.2.26+ (22) and SequenceServer (23) or prfectBLAST (24). Automatic primer design was carried out using the Primer3Plus web server (25). DoBISCUIT (http://www.bio.nite.go.jp/pks/top) (Database of Biosynthesis clusters Curated and Integrated) (26) was used to obtain curated and annotated versions of ansamycin-related gene clusters. The sequence of phage P1-derived artificial chromosome (PAC) clone 2-11L (named pIJ12853), selected from a library of *S. leeuwenhoekii* genomic DNA made in pESAC13 (*E. coli-Streptomyces* artificial chromosome [27]; http://www.biost.com/page/PAC.aspx) and containing the chaxamycin biosynthesis gene cluster (see "PAC library construction and PCR screening for the chaxamycin gene cluster" below) was obtained using a Roche 454-Junior sequencer (commissioned from the DNA sequencing facility in the Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom). A high-quality sequence of the complete PAC insert was obtained by mapping the assembled contigs from the 454-Junior sequencing onto the genome sequence (19) using the Burrows Wheeler Aligner BWA-MEM algorithm with default parameters (28); the resulting BAM alignment file was processed with SAMTOOLS (29) and loaded into GAP5 (30) to manually produce a consensus sequence. This consensus was processed with Prodigal (31) to identify protein-coding sequences, automatically annotated using the BASys (32) web server (https://www.basys.ca/), and

TABLE 3 Oligonucleotides used in this study[a]

Name	Sequence (5′ → 3′)[b]	Use
JFC009	TTTCTAGAGAAGCCATCAGCCAGGACT	Amplify upstream flanking region of AHBA synthase gene
JFC010	TTAAGCTTGTACGTCCAGCAGCTCCAGT	
JFC011	TTTCTAGAGCTCGGTCTCGTCGTACTG	Amplify downstream flanking region of AHBA synthase gene
JFC012	TTGGATCCACTGCTCCTGCTCGACTACG	
JFC022	CGATTCGAAGAGTGACAGCA	Screening of genomic library for chaxamycin gene cluster
JFC023	GTTGTAGGTGGCGATTTTGC	
JFC024	GCGAAGAGAAGAGCGAGGT	Screening of genomic library for chaxamycin gene cluster
JFC025	CACACATTGCTCAGATGCAC	
JFC026	AGACATATGAACGCGCGACAG	Cloning of *cxmK* in pIJ10257
JFC034	AAGCTTACGACCTGGTAG	
JFC032	GTCGTTCCCTCATTTCCTCA	Confirmation of deletion of AHBA synthase gene
JFC033	GACCTGGTAGCCGGTGTG	
JFC036	GACGAGAACACCACGAAGGT	Resolution of a frameshift within one of the PKS genes
JFC037	GACCACCGGCACTACTGG	
JFC053	AAGCAGGTCAGCAGGATCA	Verification of the sequence between *cxm24* and *cxmY*
JFC054	TTCCGGAAGATGAACGTGAC	

[a] Oligonucleotide pairs are shown in consecutive shaded or unshaded rows, and the information in the "Use" column applies to both relevant oligonucleotides.
[b] Incorporated restriction sites are underlined.

analyzed and curated with Artemis (33). Analysis of the sequence of the insert of pIJ12853 revealed that it corresponded to nucleotide (nt) positions 1151661 to 1305894 of the published *S. leeuwenhoekii* chromosome sequence (European Nucleotide Archive study accession number PRJEB8583) and contained many of the genes predicted to be required for chaxamycin biosynthesis (located between nt 1211049 and 1289829). GC Frame Plot analysis (34) performed within Artemis revealed a clear frameshift in the 454 sequence within one of the PKS genes. The corresponding region was PCR amplified using primers JFC036 and JFC037 and sequenced with the same primers by Sanger sequencing. Analysis revealed the presence of two close frameshifts (two missing Gs), and the consensus sequence was corrected accordingly (matching that obtained with Illumina MiSeq sequencing of the *S. leeuwenhoekii* genome [19]). GC Frame Plot analysis revealed possible frameshifts between approximate nucleotide positions 1214200 and 1215100. Indeed, BLASTX analyses (21) confirmed that this region had the potential to encode a type II thioesterase (see Discussion) but contained three frameshifts in a putative ancestral coding sequence that begins at nt 1215047 and ends at nt 1214242. To assess the quality of this sequence, the region was PCR amplified using primers JFC053 and JFC054 with pIJ12853 (see below) DNA as the template, and the fidelity of the sequence and the presence of the frameshifts were confirmed by Sanger sequencing using the same primers.

PAC library construction and PCR screening for the chaxamycin gene cluster. A genomic library of *S. leeuwenhoekii* DNA was constructed in pESAC13. Briefly, high-molecular-weight DNA was digested partially with BamHI and ligated with pESAC13 digested with the same enzyme, resulting in loss of pUC19 present in the original vector. The genomic library was screened for clones containing the chaxamycin biosynthesis gene cluster by PCR using two primer pairs, JFC022/JFC023 and JFC024/JFC025, designed to amplify the predicted ends of the gene cluster. One PAC clone (2-11L) was identified, named pIJ12853, and sequenced (see "DNA sequencing and bioinformatic analysis" above). Both library construction and screening were carried out by Bio S&T Inc. (Montreal, Canada).

Construction of *S. leeuwenhoekii* M1653 (Δ*cxmK*::*neo*). The AHBA synthase gene (*cxmK*) was deleted and replaced with the kanamycin resistance gene *neo* by double-crossover homologous recombination. Two DNA segments flanking *cxmK*, each about 1.7 kb in length, were PCR amplified with primer pairs JFC010/JFC009 (containing 5′ HindIII and XbaI sites, respectively) and JFC011/JFC012 (containing 5′ XbaI and BamHI sites, respectively) and cloned separately in pKC1132 that had been cleaved with HindIII plus XbaI or XbaI plus BamHI to generate pIJ12857 and pIJ12858, respectively. Each of the amplified fragments was verified by DNA sequencing. pIJ12858 was cleaved with XbaI and BamHI, and the 1.7-kb band was purified from an agarose gel and cloned into pIJ12857 cut with XbaI and BamHI to generate pIJ12859. The HindIII-BamHI insert of pIJ12859 was cloned in the self-replicative temperature-sensitive plasmid pGM1190 that had been cleaved with HindIII and BamHI to yield pIJ12860. Finally, *neo* was excised with XbaI from pTC192-km and cloned in the XbaI site of pIJ12860 to yield pIJ12850. This construct was transferred into *S. leeuwenhoekii* by conjugation, and kanamycin-resistant exconjugants were selected at 30°C, restreaked on SFM agar containing kanamycin, and cultivated at 37°C, a temperature at which pGM1190 cannot replicate, therefore forcing chromosomal integration of *neo* by either single or double homologous recombination. The resulting spores were plated on SFM agar without antibiotic selection and incubated at 30°C, and the resulting colonies were replicated consecutively onto SFM agar plates containing just kanamycin, just apramycin, or neither antibiotic, which were then incubated at 37°C. Colonies resistant to kanamycin but sensitive to apramycin were subjected to PCR analysis using primers (JFC032/JFC033), which flanked *cxmK*, and the required Δ*cxmK*::*neo* deletion mutant (M1653) was identified (see Fig. S7 in the supplemental material). Note that we resorted to using a derivative of the replicative, multicopy pGM1190 for gene replacement after attempts to use pIJ12851 (Table 2), a derivative of the nonreplicative pKC1132 containing the same deletion cassette as pIJ12850, failed to yield single-crossover recombinants, which possibly was a consequence of a relatively low recombination frequency in *S. leeuwenhoekii*. We hoped to compensate for any such deficiency by using the multicopy, replicative pGM1190 derivative.

Construction of pIJ12852. *cxmK* was amplified from *S. leeuwenhoekii* genomic DNA by high-fidelity PCR using primers JFC026 and JFC034 (containing 5′ NdeI and HindIII sites, respectively); the PCR product was blunt-end ligated into the SmaI site of pBluescript II KS(+) to yield pIJ12861, and the sequence of the PCR product was verified by sequencing with universal primers corresponding to the vector. *cxmK* was excised from pIJ12861 with NdeI and HindIII and inserted into pIJ10257 that had been cleaved with the same enzymes to generate pIJ12852 with *cxmK* transcribed from the constitutive *ermE** promoter.

Generation of illustrations. Illustrations of LC-MS/MS data were generated using Microsoft Office Suite. ImageJ version 1.47 (http://imagej.nih.gov/ij) and Inkscape version 0.91 (https://inkscape.org) were used to create and edit drawings and illustrations.

RESULTS

Establishment of microbiological and genetic manipulation procedures for *S. leeuwenhoekii*. *S. leeuwenhoekii* grew well on SFM, R2, and DNA agar media, with the most copious and rapid sporulation observed after growth for 4 to 6 days at 30°C on SFM agar (see Fig. S1 in the supplemental material). Since *S. leeuwenhoekii* was isolated from the Atacama Desert (Chile), we compared growth at 30°C, 37°C, and 43°C. Growth and sporulation occurred more rapidly with increasing temperature, with sporulation occurring at 37°C and 43°C after just 2 days of cultivation (see Fig. S2 in the supplemental material). Consequently, we used 37°C to prepare spore stocks, a compromise between high growth rate and sporulation and a temperature at which the medium did not dry out too quickly. Chaxamycin production was also assessed at the same three temperatures in modified ISP2 production medium and shown to be very low at 37°C and 43°C compared to 30°C (see Fig. S3 in the supplemental material); consequently, all studies of chaxamycin production were performed at 30°C.

Before establishing techniques for the genetic manipulation of *S. leeuwenhoekii*, we first tested the susceptibility of the strain to a range of antibiotics commonly used with streptomycetes (16). Growth was prevented at concentrations of >1.0 μg/ml kanamycin, >0.5 μg/ml apramycin, >2.0 μg/ml hygromycin B, and >2.5 μg/ml thiostrepton; *S. leeuwenhoekii* was not sensitive to phosphomycin up to 50 μg/ml, while growth was retarded in the presence of nalidixic acid at >25 μg/ml (see Fig. S4 in the supplemental material).

Many plasmids used to manipulate streptomycetes genetically utilize the attachment sites of temperate phages (16). Putative bacterial attachment (*attB*) sites for the phiC31 and phiBT1 actinophages were identified in the chromosome of *S. leeuwenhoekii* (European Nucleotide Archive accession number PRJEB8583 [19]) by BLAST searches using the sequences of the corresponding sites from *S. coelicolor* (35, 36) (see Table S1 in the supplemental material). In both cases, the putative *attB* sites were located in homologues of the genes in which they reside in *S. coelicolor* (see Table S2 in the supplemental material).

Conditions for transferring DNA into *S. leeuwenhoekii* by conjugation were established using pSET152, which integrates in the phiC31 *attB* and carries the apramycin resistance gene *aac(3)IV*, and pIJ10257, which integrates at the phiBT1 *attB* and carries the

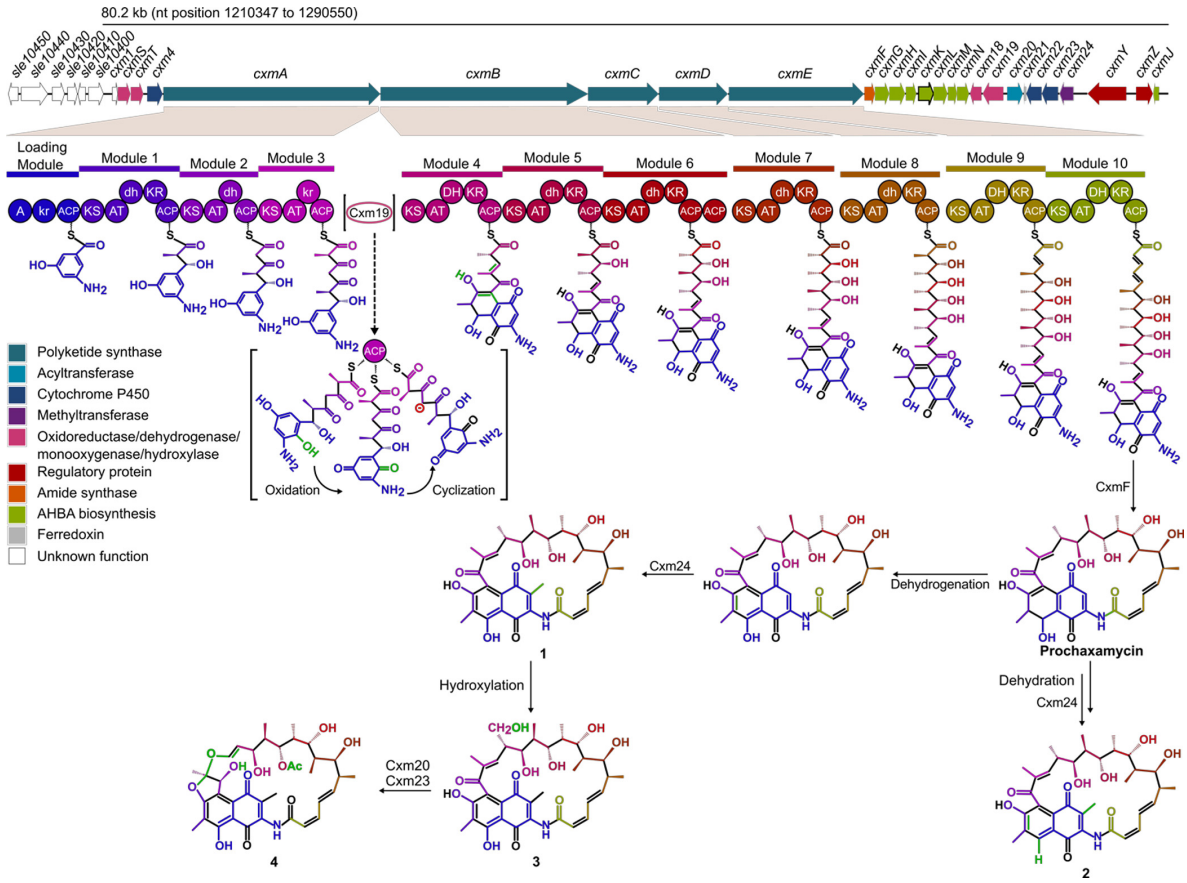


FIG 2 Top, organization of the chaxamycin gene cluster (*cxm*) of *S. leeuwenhoekii*. The proposed function of each gene is listed in Table 2. Bottom, proposed pathway for chaxamycin biosynthesis (based on that for rifamycin). The modules encoded in the polyketide synthase genes and their respective domains are as follows: A, adenylation; KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; DH, dehydratase; ACP, acyl carrier protein. The domains in lower case should be inactive. Chaxamycins: 1, chaxamycin A; 2, chaxamycin B; 3, chaxamycin C; 4, chaxamycin D.

hygromycin B resistance gene *hyg*; conjugations were plated on SFM agar supplemented with 10 mM $MgCl_2$ and 10 mM $CaCl_2$.

The most common constitutive promoter used for gene expression in *Streptomyces* is *ermE**p (37). To test its usability in *S. leeuwenhoekii*, we introduced pGUS (a pSET152 derivative with the *gusA* reporter gene [38]) and pIJ10740 (a pGUS derivative with *ermE**p driving the transcription of *gusA* [Morgan Feeney, personal communication]); *gusA* expression (assessed as described previously [39]) was readily observed in the pIJ10740 exconjugants but not in the pGUS exconjugants or in the *S. leeuwenhoekii* wild-type strain, confirming the utility of *ermE**p in this species.

pGM1190 is a self-replicative, temperature-sensitive plasmid (reported to be nonreplicative at over 34°C) used for gene replacement in *Streptomyces* (40). We readily obtained *S. leeuwenhoekii* exconjugants with pGM1190 and verified plasmid loss after cultivation at 34°C, 37°C, and 43°C (see Fig. S5 in the supplemental material). Since we observed some persistence of pGM1190 at 34°C, 37°C was used to cure the plasmid.

Bioinformatic identification of the chaxamycin biosynthesis gene cluster. Bioinformatic analysis of the genome sequence of *S. leeuwenhoekii* identified a contiguous stretch of DNA that contained many of the genes predicted to be required for chaxamycin biosynthesis, including those needed for AHBA biosynthesis (i.e., the previously sequenced AHBA synthase gene [GenBank accession number FR839674.1]) and a complete set of appropriate PKS modules, including a loading module with similarity to those of other ansamycin-type PKSs such as ansamitocin (DoBISCUIT accession code Ansam_00270), geldanamycin (Gelda2_00080), herbimycin A (Herb_00170), rifamycin (Rifam_00210), and rubradirin (Rubra_00070).

The putative limits of the gene cluster were identified by comparison with gene clusters encoding other ansamycin-like polyketides, including rifamycin (the *rif* cluster) from *A. mediterranei* S699 (AF040570.3), naphthomycin from *Streptomyces* sp. strain CS (GQ452266.1), and saliniketal from *Salinispora arenicola* CNS-205 (CP000850.1). The right end of the chaxamycin gene cluster (as shown in Fig. 2 and in Fig. S6 in the supplemental material) was defined by the presence of a gene, *cxmJ*, homologous to *rifJ*, encoding a putative aminodehydroquinate (amino-DHQ) dehydratase involved in the biosynthesis of AHBA (catalyzing the

TABLE 4 Proposed chaxamycin biosynthesis gene cluster and gene functions

sle no.	*cxm* name[a]	No. of amino acids	Rifamycin homologue (mutant phenotype)[b]	Proposed function in chaxamycin biosynthesis[c]
sle10390	*cxm1*	138		Small hypothetical protein
sle10380	*cxmS*	327	*rifS*	NADH-dependent oxidoreductase
sle10370	*cxmT*	323	*rifT*	NADH-dependent dehydrogenase
sle10360	*cxm4*	397	*orf0*	Cytochrome P450
sle10350	*cxmA*	5616	*rifA*	Polyketide synthase (module 0 [M0], A_{AHBA}-KR-ACP; M1, KS-AT_{mmal}-dh-KR-ACP; M2, KS-AT_{mal}-DH-ACP; M3, KS-AT_{mmal}-KR-ACP)
sle10340	*cxmB*	5363	*rifB*	Polyketide synthase (M4, KS-AT_{mmal}-dh-KR-ACP; M5, KS-AT_{mmal}-dh-KR-ACP; M6, KS-AT_{mmal}-dh-KR-ACP-ACP)
sle10330	*cxmC*	1820	*rifC*	Polyketide synthase (M7, KS-AT_{mmal}-dh-KR-ACP)
sle10320	*cxmD*	1773	*rifD*	Polyketide synthase (M8, KS-AT_{mmal}-DH-KR-ACP)
sle10300	*cxmE*	3488	*rifE*	Polyketide synthase (M9, KS-AT_{mal}-DH-KR-ACP; M10, KS-AT_{mal}-DH-KR-ACP)
sle10290	*cxmF*	269	*rifF* (−)[e]	Polyketide release and ansa-ring formation: amide synthase
sle10280	*cxmG*	368	*rifG*[d]	AHBA synthesis, aminodehydroquinate synthase
sle10270	*cxmH*	406	*rifH*[d]	AHBA synthesis, amino-DAHP synthase
sle10260	*cxmI*	268	*rifI*[d]	AHBA synthesis, aminoquinate dehydrogenase
sle10250	*cxmK*	386	*rifK*[d]	AHBA synthesis, AHBA synthase
sle10240	*cxmL*	358	*rifL*[d]	AHBA synthesis, oxidoreductase
sle10230	*cxmM*	232	*rifM*[d]	AHBA synthesis, phosphatase
sle10220	*cxmN*	307	*rifN*[d]	AHBA synthesis, kanosamine kinase
sle10210	*cxm18*	295	*orf11* (+)[f]	Flavin-dependent oxidoreductase
sle10200	*cxm19*	533	*orf19* (−)[f]	Naphthalene ring formation, FDA-dependent-monooxygenase and 3-(3-hydroxylphenyl) propionate hydroxylase
sle10190	*cxm20*	402	*orf20*	Tailoring, *O*-acyltransferase (chaxamycin D)
sle10180	*cxm21*	63		Ferredoxin
sle10170	*cxm22*	393	*orf4* (+)[f]	Cytochrome P450 monooxygenase
sle10160	*cxm23*	418	*orf5* (W)[f], *orf13* (+)[f]	Tailoring, cytochrome P450 monooxygenase (hydroxyfuran of chaxamycin D)
sle10150	*cxm24*	355		Tailoring: methyltransferase, *S*-adenosylmethionine dependent (not homologue of *orf14*)
sle10120	*cxmY*	433		Transcriptional regulator (C-terminal DNA-binding domain found in the NarL/FixJ response regulator family)
sle10110	*cxmZ*	246		Transcriptional regulator, atypical response regulator of OmpR/PhoB family
sle10100	*cxmJ*	168	*rifJ*[d]	AHBA synthesis, aminodehydroquinate dehydratase

[a] Gene names have been assigned as far as possible according to predicted functional homology to the rifamycin gene cluster (see also Table S3 in the supplemental material).
[b] Rifamycin gene cluster from *A. mediterranei* S699 (GenBank accession no. AF040570.3). Mutant phenotypes: +, similar production to parent strains; −, production abolished; W, loss of rifamycin B production and accumulation of rifamycin W.
[c] Based on Pfam motif search and homology to rifamycin biosynthesis proteins. PKS domains (found with antiSMASH [47] and NCBI-CDD [72]): A, adenylation; ACP, acyl carrier protein; AT, acyltransferase; DH, dehydratase (lowercase indicates that it should be inactive); KR, ketoreductase; KS, ketosynthase. The specificity predicted for each AT domain is shown as a subscript: AHBA, 3-amino-5-hydroxybenzoic acid; mal, malonyl-CoA; mmal, methylmalonyl-CoA.
[d] Genes required for AHBA biosynthesis (43).
[e] According to reference 56.
[f] Mutant accumulates linear polyketide intermediates (59).

conversion of 5-deoxy-5-aminodehydroquinic acid to 5-deoxy-5-aminodehydroshikimic acid [amino-DHS] [14]); surprisingly, *cxmJ* is located some distance from the other genes required for AHBA biosynthesis, as also noted for the *cxmJ* homologues in the other three gene clusters (13).

The left end of the chaxamycin gene cluster (as shown in Fig. 2) could not be readily predicted, although from comparisons with the other ansamycin-like gene clusters, we initially speculated that *cxmS*, encoding a putative NADH-dependent dehydrogenase homologous to the *rifS* product, would be the likely limit of the cluster. At 8.7 kb upstream of *cxmS* (to the left of *cxmS* in Fig. 2), we found a conserved region of 11 genes (*sle10460* to *sle10560*) including a putative hopanoid biosynthesis gene cluster that is also present in *S. scabies* 87.22 (*scab_12881* to *scab_13001*, excluding *scab_12921* [41]) and *S. coelicolor* (*sco6759* to *-6771*, except *sco6761* [42]), as well as in many other streptomycetes. Between *cxmS* and the hopanoid biosynthesis gene cluster there are seven genes, four of which encode highly conserved proteins with homologues in *S. coelicolor*: Sle10450, Sle10430, Sle10420, and Sle10400 are homologous to Sco6772, Sco6773, Sco6774, and Sco6776, respectively. Two of the remaining genes (*sle10440* and *sle10410*) encode highly conserved proteins that have no close homologues in *S. coelicolor*. The gene immediately upstream of *cxmS*, *sle10390*, encodes a hypothetical protein without clear homologues in the databases; since the 3′ end of the coding sequence of *sle10390* overlaps with *cxmS* and there is a 720-bp intergenic region between *sle10390* and the gene immediately upstream, *sle10400*, it is conceivable that *sle10390* is part of the chaxamycin biosynthesis gene cluster, and thus this gene was assigned as the left limit of the cluster and named *cxm1* (Fig. 2 and Table 4).

The chaxamycin biosynthesis gene cluster of *S. leeuwenhoekii*, as defined above, spans a region of 80.2 kb, from nt 1210347 to 1290550 of the published genome sequence (19), and contains 27 genes. The core PKS and AHBA biosynthesis genes show very

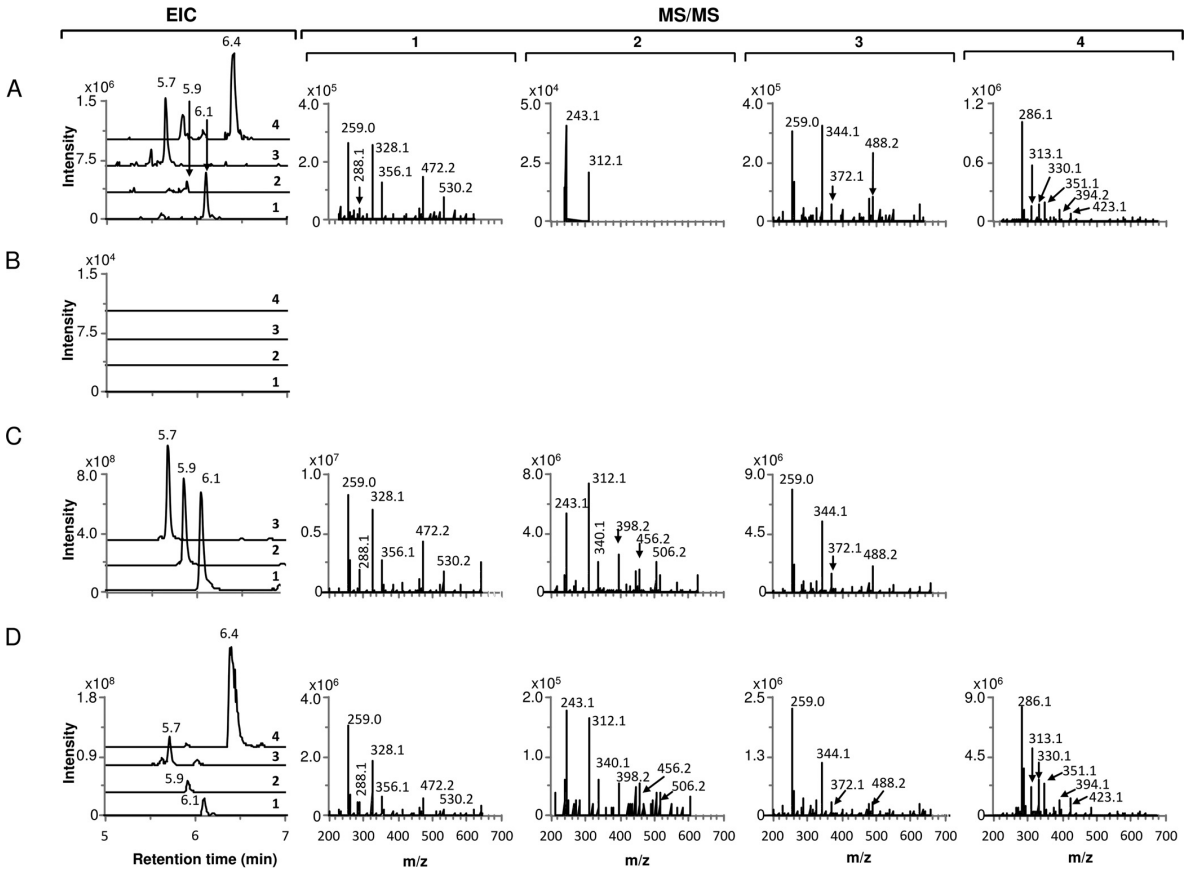


FIG 3 Heterologous production of chaxamycins A to D in *S. coelicolor* M1152. Extracted ion chromatogram (EIC) and MS/MS fragmentation patterns are shown for each chaxamycin species detected. (A) *S. coelicolor* M1650 (M1152 containing pIJ12853); (B) *S. coelicolor* M1152 (negative control); (C) chaxamycin A, B, and C standards; (D) *S. leeuwenhoekii*. Chaxamycin A (compound 1), *m/z* $(M - H)^-$ 638.29; chaxamycin B (compound 2), *m/z* $(M - H)^-$ 622.29; chaxamycin C (compound 3), *m/z* $(M - H)^-$ 654.29; chaxamycin D (compound 4), *m/z* $(M - H)^-$ 682.29.

similar organization, and the encoded proteins high sequence identity, to the rifamycin biosynthesis gene cluster of *A. mediterranei* S699 (GenBank accession no. AF040570.3) (see Fig. S6 in the supplemental material). Consequently, gene names and predicted functions were assigned based on those defined in the rifamycin gene cluster (Fig. 2 and Table 4).

Heterologous expression of the chaxamycin biosynthesis gene cluster. One clone, pIJ12853, from a PAC library of *S. leeuwenhoekii* genomic DNA was identified by PCR and confirmed by sequencing to contain the proposed chaxamycin biosynthesis gene cluster (see Materials and Methods). pIJ12853 was mobilized into *S. coelicolor* M1152, resulting in integration of the gene cluster at the chromosomal phiC31 *attB* site, yielding strain M1650.

M1650 was cultivated in R3 liquid medium, and culture supernatants and extracts of mycelial samples harvested at different times were analyzed by LC-MS, in parallel with chaxamycin standards and samples from *S. leeuwenhoekii* grown in modified ISP2 liquid medium. The expected molecular ions for chaxamycins A to D (compounds 1 to 4) were readily found in samples from M1650 and with retention times (6.1, 5.9, 5.7, and 6.4 min, respectively) similar to those for the purified standards and the samples from *S. leeuwenhoekii* (Fig. 3). In addition, the MS/MS fragmentation patterns of the ions detected for compounds 1, 3, and 4 in the M1650 samples matched those detected from the natural producer, while those for compounds 1 and 3 also matched the fragmentation patterns obtained from the standards; production of compound 2 in M1650 was very low and did not provide clear fragmentation data (Fig. 3B). These results demonstrated that the proposed gene cluster does indeed encode chaxamycin biosynthesis.

Deletion of the AHBA synthase gene (*cxmK*) and chemical complementation. AHBA is the starter unit for the synthesis of the polyketide backbone of chaxamycin. *cxmK* (Table 4) encodes a putative AHBA synthase that catalyzes both the transamination of UDP-3-keto-α-D-glucose at an early stage of the pathway and the final aromatization of amino-DHS to AHBA (13). Only one copy of this gene was found in the genome sequence of *S. leeuwenhoekii*, directly downstream of the PKS genes. *cxmK* was deleted and replaced with the kanamycin resistance gene *neo* by homologous recombination (see Materials and Methods and Fig. S7 in the sup-

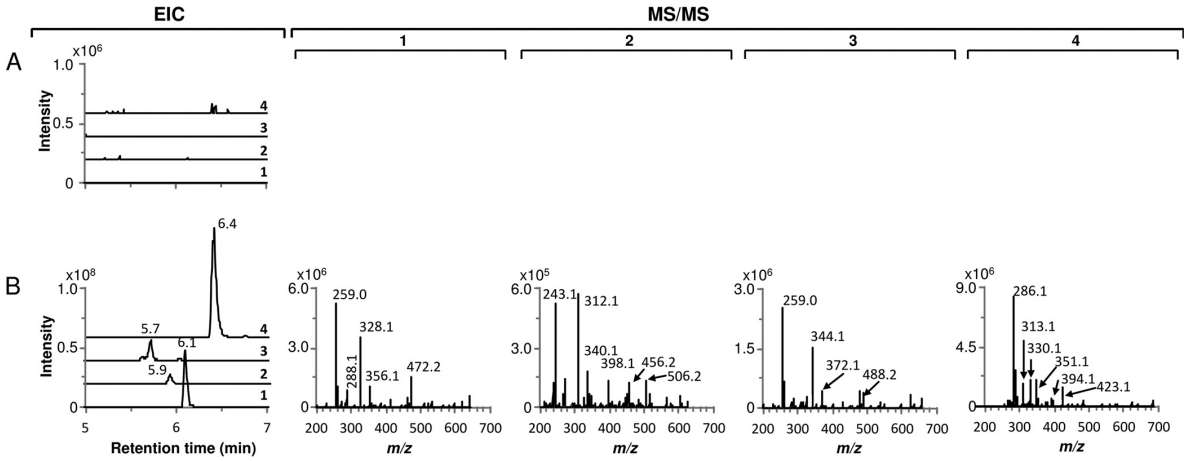


FIG 4 Chemical complementation of *S. leeuwenhoekii* M1653 (Δ*cxmK*::*neo*) with 3-amino-5-hydroxybenzoic acid (AHBA). Extracted ion chromatogram (EIC) and MS/MS fragmentation patterns are shown for each chaxamycin species detected. (A) *S. leeuwenhoekii* M1653; (B) *S. leeuwenhoekii* M1653 supplemented with 0.36 mM AHBA. Chaxamycin A (compound 1), *m/z* $(M - H)^-$ 638.29; chaxamycin B (compound 2), *m/z* $(M - H)^-$ 622.29; chaxamycin C (compound 3), *m/z* $(M - H)^-$ 654.29; chaxamycin D (compound 4), *m/z* $(M - H)^-$ 682.29.

plemental material) to generate *S. leeuwenhoekii* M1653 (Δ*cxmK*::*neo*).

M1653 was cultivated in parallel with wild-type *S. leeuwenhoekii* in chaxamycin production medium; LC-MS analysis of samples of culture supernatants harvested at different times after inoculation showed that biosynthesis of all of the chaxamycin species had been abolished in the Δ*cxmK* mutant M1653 (Fig. 4A). To confirm that the lack of production was due solely to the deletion of *cxmK*, we attempted to complement the mutant by expression in *trans* of a wild-type copy of *cxmK* cloned in the expression vector pIJ10257. The resulting plasmid, pIJ12852, was introduced into M1653 but failed to restore chaxamycin production, presumably reflecting a polar effect of the insertion of *neo* on the expression of other downstream AHBA biosynthesis genes. We then succeeded in complementing the mutant chemically, restoring chaxamycin production (Fig. 4B) by supplementing the culture medium with commercially available AHBA; after testing several concentrations and times of addition, optimal results were obtained when adding AHBA to 0.36 mM at 24 h after inoculation, presumably coinciding with production of the chaxamycin biosynthesis enzymes and minimizing AHBA degradation and/or catabolism. These feeding experiments confirmed a role for *cxmK* in AHBA synthesis and provided further evidence that the identified gene cluster does indeed encode chaxamycin biosynthesis. It also suggests that mutasynthesis, and the feeding of AHBA analogues, may be a viable approach for the generation of unnatural chaxamycin variants.

DISCUSSION

Functions of the genes present in the chaxamycin biosynthesis gene cluster and a proposed biosynthetic pathway. We have unequivocally identified, cloned, genetically manipulated, and heterologously expressed the chaxamycin biosynthesis gene cluster from *S. leeuwenhoekii*. The chemical structures of the chaxamycins and the rifamycins share many similarities, and therefore it is not surprising that the biosynthesis gene clusters are also very similar in organization. Based on knowledge of rifamycin biosynthesis, we can now propose a biosynthetic pathway for chaxamycin.

The chaxamycin biosynthesis gene cluster (Fig. 2 and Table 4) contains homologues (*cxmG* to *-N* and *cxmJ*) of eight of the genes involved in AHBA biosynthesis in the rifamycin gene cluster of *A. mediterranei* S699 (*rifG* to *-N* and *rifJ*; like *rifJ*, *cxmJ* is not clustered with the other genes involved in AHBA synthesis). They show the same organization as the *rif* genes, and their products show high levels of amino acid sequence identity (68 to 86%). Therefore, we propose that biosynthesis of the chaxamycin PKS starting unit (AHBA) occurs as in rifamycin biosynthesis in *A. mediterranei* S699 (see Fig. S8 in the supplemental material) (13, 14, 43).

In addition to CxmG to -N and CxmJ, a transketolase is also required for AHBA biosynthesis to catalyze the conversion of 3-amino-3-deoxy-D-fructose-6-phosphate to imino-erythrose-4-phosphate (14). Indeed, in *Salinispora arenicola* CNS-205, deletion of *sare_1272* and *sare_1273*, which encode a two-subunit transketolase, from the shared rifamycin-saliniketal biosynthesis gene cluster abolished the production of both compounds (44). While the proposed chaxamycin biosynthesis gene cluster does not appear to possess such a function, there are two genes (*sle09700* and *sle09710*) located 47.4 kb to the right of it, contained within the insert of pIJ12853, that encode homologues of transketolase subunits from *A. mediterranei* (encoded by *rif15A* and *rif15B*, respectively [45]) that might fulfil this role. Other lines of evidence suggest that such a function can be provided by genes outside the biosynthesis gene cluster. In *A. mediterranei*, deletion of *rif15A* and *rif15B* (45), homologues of *sare_1272* and *sare_1273*, respectively, led to the accumulation of rifamycins SV and S, indicating that AHBA was still being produced in the mutant strain. In addition, the heterologous production of AHBA in *S. coelicolor* after cloning the AHBA genes *rifG* to *-N* and *rifJ* suggests that the transketolase step can be performed by an alternative enzyme encoded elsewhere in the genome (14, 43).

CxmI is homologous to RifI, an aminoquinate (amino-Q) dehydratase that catalyzes the conversion of amino-DHQ to amino-Q, with no obvious role in AHBA biosynthesis. Deletion of *rifI* in *A. mediterranei* led to the accumulation of 20 to 25% more AHBA, and a salvage function has been proposed for the enzyme (43, 46).

The type I PKS, responsible for the assembly of the chaxamycin polyketide backbone, is encoded by five genes, *cxmA* to *-E*. Bioinformatic analysis of their protein sequences was performed with antiSMASH (47), which uses two methods to predict the substrate specificity of the loading and AT domains (48, 49). antiSMASH predicted a loading module with specificity for AHBA, followed by a module selective for methylmalonyl-CoA, one for malonyl-CoA, six more for methylmalonyl-CoA, and finally two more for malonyl-CoA, which is entirely consistent and colinear with the reactions required to produce chaxamycin A (compound 1) (Fig. 2 and Table 4).

The loading module of CxmA contains an adenylation domain with 79% sequence identity to the adenylation domain of RifA in *A. mediterranei*, which catalyzes the reaction between a benzoate molecule and ATP to produce an activated aryl-AMP species (50, 51). The CxmA adenylation domain contains the AMP-binding and ATPase motifs, suggesting that it could function in a similar fashion to its homologue from *A. mediterranei* (see Fig. S9 in the supplemental material).

The chemical structure of compound 1 suggests that the dehydratase (DH) domains from modules 1, 2, 5, 6, 7, and 8 should be inactive. Active DH domains possess two conserved motifs, ^{2409}HXXXGXXXXP and ^{2571}D(A/V)(V/A)(A/L)(Q/H), where H2409 and D2571 (the numbers refer to amino acid positions in the DH4 domain of the erythromycin PKS of *Saccharopolyspora erythraea*) comprise the catalytic dyad of the dehydratase (52, 53). DH domains from modules 4, 9, and 10 of the chaxamycin PKS, predicted to be functional, all contain the second conserved motif DAALH and the conserved histidine residue in the motif HXXX GXXXXP (although the DH domain from module 9 has proline replaced by alanine). DH domains from modules 1 and 2, predicted to be nonfunctional, do not have any of the conserved motifs. DH domains from modules 5 and 8, also predicted to be nonfunctional, are shorter than the rest of the DH domains and do not contain the motif D(A/V)(V/A)(A/L)(Q/H); thus, despite having the conserved H residue in the active site, they lack the catalytic dyad. Surprisingly, the DH domains of modules 6 and 7 contain both active-site motifs, and we presume that there are other mutations that result in their inactivity (see Fig. S10 in the supplemental material).

The KR domain of module 3 lacks the active-site tyrosine residue (54) and the NADPH-binding motifs (55); the KR domain present in the loading module is shorter and lacks both of sequence features mentioned above. Therefore, these KR domains are predicted to be inactive, again in agreement with the structure of chaxamycin. The remaining KR domains have the tyrosine residue and the conserved motifs for NADPH binding (see Fig. S11 in the supplemental material) and are predicted to be active, consistent with the structure of the polyketide (Fig. 2).

The chaxamycin PKS shows a high level of similarity to that involved in rifamycin biosynthesis in *A. mediterranei* S699 but contains two additional nonfunctional domains (an extra KR domain in the loading module and an extra DH domain in module 2) as well as, unusually, an apparently functional additional acyl carrier protein (ACP) domain in module 6. As for the rifamycin PKS, the chaxamycin PKS lacks a type I thioesterase domain at the end of module 10, and a putative *N*-acetyltransferase/amide synthase is encoded by *cxmF*, the likely orthologue of *rifF*, and is similarly located just downstream of the last PKS gene (*cxmE*) (Fig. 2). Therefore, we propose that CxmF catalyzes the formation of the internal amide bond between AHBA and the last extender molecule of the polyketide chain to close the macrolactam ring, as in rifamycin biosynthesis (56), thus releasing the PKS product, which we have called prochaxamycin (Fig. 2).

In rifamycin biosynthesis, a type II thioesterase, RifR, is thought to function in the release of aberrant intermediates from the PKS, thus restoring functionality for a new cycle (57). However, it is not essential, and the PKS itself may be capable of such activity (58). Interestingly, we identified a nonfunctional *rifR* homologue located between *cxmY* and *cxm24* that contained three frameshift mutations (the frameshift-corrected amino acid sequence shows end-to-end similarity and 64% identity to RifR); possible errors in the PacBio-Illumina assembled DNA sequence were ruled out by PCR amplification and Sanger sequencing (see Materials and Methods). Whether an alternative type II thioesterase encoded outside the *cxm* gene cluster plays an editing role for the chaxamycin polyketide synthase remains to be determined.

We identified eight genes encoding putative tailoring activities homologous to those found in the rifamycin gene cluster of *A. mediterranei*. In rifamycin biosynthesis, the naphthalene ring is formed during the extension of the polyketide backbone by the enzyme 3-(3-hydroxylphenyl)propionate hydroxylase encoded by *orf19* (59); a homologous gene, *cxm19*, is also present in the chaxamycin gene cluster, and the encoded protein is predicted to catalyze the formation of the naphthalene ring of chaxamycin.

A putative methyltransferase is encoded by *cxm24*, but it shows little similarity (25% identity over 87 residues of the 355-amino-acid Cxm24) to the C-27 *O*-methyltransferase (Rif-Orf14) involved in rifamycin biosynthesis. This is the only putative methyltransferase encoded by the chaxamycin gene cluster, and although Pfam predicts that it is an *O*-methyltransferase, we speculate that it is responsible for C methylation of the C-3 of AHBA in the naphthalene ring.

An *O*-acetyltransferase encoded by *orf20* catalyzes the O acetylation of C-25 during rifamycin biosynthesis (60). *cxm20*, encoding a homologous protein with 63% identity to Rif-Orf20, is thus proposed to catalyze the O acetylation of the hydroxyl group of C-25 in chaxamycin D (9). In rifamycin biosynthesis by *A. mediterranei* S699, a cytochrome P450 monooxygenase (CYP) encoded by *orf5* is required for the formation of the ketofuran moiety present in rifamycins SV and B but not in rifamycin W; a very similar CYP encoded by *orf13* seems to be nonessential for rifamycin B biosynthesis (59). We identified only one gene, *cxm23*, which encodes a putative CYP homologous to rifamycin Orf5 or Orf13 (68% and 72% amino acid sequence identity, respectively). This is similar to the rifamycin gene cluster from *S. arenicola* CNS-205, where only one encoded protein, Sare1259, is homologous to Orf5 and Orf13 from *A. mediterranei* S699 (with higher identity to Orf13 [79%] than to Orf5 [71%]). Sare1259 performs several oxidation reactions at different positions on the molecule (44) and is 77% identical to Cxm23; therefore, we propose that Cxm23 participates in the formation of the hydroxyfuran ring present in chaxamycin D but not in chaxamycin A, B, or C (Fig. 2).

Transcriptional regulation. Two genes encode putative tran-

scriptional regulatory proteins, *cxmY* and *cxmZ*. Analysis of CxmY with Pfam (61) revealed a carboxy-terminal helix-turn-helix (HTH) DNA-binding domain (PF00196) typical of the NarL/FixJ family of response regulators (RRs) (62); the rest of the amino acid sequence shows no similarity to that protein family, but BLASTp analysis (21) revealed many homologues of putative transcriptional regulators in other actinomycetes, including Rif-Orf36 of the rifamycin gene cluster of *A. mediterranei* S699, with which it shows end-to-end similarity and 43% amino acid sequence identity. Similar analysis of CxmZ revealed a carboxy-terminal DNA-binding domain of the winged HTH family (Trans_reg_C family, PF00486) typical of the OmpR/PhoB family of RRs (62), and an amino-terminal RR receiver domain (Response_reg family, PF00072). CxmZ lacks the first of a pair of aspartates at positions 13 (serine) and 14 (aspartate) that determine the structure of the phosphorylation pocket in typical RRs, and it is thus unlikely to be phosphorylated by a cognate histidine sensory kinase (no such protein is encoded in the *cxm* gene cluster), supporting its classification as an atypical RR (see Fig. S12 in the supplemental material). CxmZ shows 41% amino acid sequence identity to JadR1, an atypical response regulator involved in the regulation of jadomycin biosynthesis in *Streptomyces venezuelae* (GenBank accession no. AAB36584.2). Although contiguous, *cxmY* and *cxmZ* are transcribed in the opposite direction with an intergenic region of over 600 bp that presumably contains the promoter sequences of both genes.

Transport and immunity. Surprisingly we did not find any gene(s) encoding a putative transport system within, or in close proximity to, the *cxm* gene cluster. The rifamycin biosynthesis gene cluster of *A. mediterranei* S699 encodes an efflux transporter (encoded by *rifP*) whose transcription is likely repressed by a TetR-like transcriptional regulator encoded by the gene immediately downstream (*rifQ*) (46, 63). Homologues of these two genes are absent from the chaxamycin gene cluster and from the rifamycin/saliniketal gene cluster of *S. arenicola* CNS-205, which also does not seem to encode a transport system (44). However, two RifP homologues are encoded elsewhere in the *S. leeuwenhoekii* genome: Sle37020 and Sle08130 (62% and 60% amino acid sequence identity to RifP, respectively). Neither of the corresponding genes is present in pIJ12853. *S. coelicolor* M145, from which strain M1152 was derived, also encodes a RifP homologue, Sco4024 (62% identity to RifP). Sco4024 is 78% identical to Sle37020 and 64% identical to Sle08130. All of these predicted proteins (as well as RifP) contain the major facilitator superfamily (MFS) domain and could be responsible for export of chaxamycin from their respective strains. A *rifP* homologue (*sare_2597*) also occurs in the genome of *S. arenicola* CNS-205, outside the rifamycin/saliniketal gene cluster.

Rifamycin binds to the β subunit (RpoB) of RNA polymerase, and in *A. mediterranei* a rifamycin-resistant form of the protein is encoded by an *rpoB* homologue located immediately downstream of *rifJ-orf36-orf37* (64). In *S. leeuwenhoekii*, *rpoB* (nt 3577583 and 3581068) lies 1.29 Mb away from the chaxamycin gene cluster (nt 1211049 to 1289829), and its product does not have the amino acid residues required for resistance to rifamycin at the expected positions, instead possessing the same amino acids as *S. coelicolor* M145 RpoB, which is sensitive to rifamycin (see Fig. S13 in the supplemental material). Thus, it seems unlikely that chaxamycin exerts its antibiotic activity by binding to RpoB.

This study has provided new insights into chaxamycin biosynthesis. We have succeeded in expressing the gene cluster heterologously, thus providing a convenient means for a more detailed study of the biosynthetic pathway and the enzymes involved. We have also shown that genetic manipulation of the natural producer is feasible, which could aid in the future engineering of the strain and the directed production of particular chaxamycin species.

ACKNOWLEDGMENTS

We are grateful to Michael Goodfellow and Alan Bull for providing *S. leeuwenhoekii*, to Marcel Jaspars and Mostafa Rateb for providing chaxamycin A, B, and C standards, and to Andriy Luzhetskyy, Morgan Feeney, Günther Muth, David Figurski, and Margherita Sosio for provision of plasmids. We thank Marcel Jaspars and Andrew Truman for comments on the manuscript.

J.F.C. and V.R. received National Ph.D. Scholarships (21110356 and 21110384, respectively) and Visiting Student Scholarships (Becas Chile, 2013-2014) from the National Commission for Scientific and Technological Research (CONICYT). This work was supported financially by the Biotechnological and Biological Sciences Research Council (BBSRC) (United Kingdom) Institute Strategic Programme Grant "Understanding and Exploiting Plant and Microbial Secondary Metabolism" (BB/J004561/1) and the Basal Programme of CONICYT (Chile) for funding of the Centre for Biotechnology and Bioengineering (CeBiB) (project FB0001).

REFERENCES

1. **Davies SC.** 2013. Annual report of the Chief Medical Officer, vol 2, 2011, Infections and the rise of antimicrobial resistance. Department of Health, London, United Kingdom.
2. **Centers for Disease Control and Prevention.** 2013. Antimicrobial resistance threats in the United States, 2013. Centers for Disease Control and Prevention, Atlanta, GA.
3. **World Health Organization.** 2014. Antimicrobial resistance: global report on surveillance. World Health Organization, Geneva, Switzerland.
4. **Baltz RH.** 2008. Renaissance in antibacterial discovery from actinomycetes. Curr Opin Pharmacol **8:**557–563. http://dx.doi.org/10.1016/j.coph.2008.04.008.
5. **Shoichet BK.** 2013. Drug discovery: nature's pieces. Nat Chem **5:**9–10. http://dx.doi.org/10.1038/nchem.1537.
6. **Wright GD.** 2014. Something old, something new: revisiting natural products in antibiotic drug discovery. Can J Microbiol **60:**147–154. http://dx.doi.org/10.1139/cjm-2014-0063.
7. **Okoro CK, Brown R, Jones AL, Andrews BA, Asenjo JA, Goodfellow M, Bull AT.** 2009. Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. Antonie Van Leeuwenhoek **95:**121–133. http://dx.doi.org/10.1007/s10482-008-9295-2.
8. **Bull AT, Asenjo JA.** 2013. Microbiology of hyper-arid environments: recent insights from the Atacama Desert, Chile. Antonie Van Leeuwenhoek **103:**1173–1179. http://dx.doi.org/10.1007/s10482-013-9911-7.
9. **Rateb ME, Houssen WE, Arnold M, Abdelrahman MH, Deng H, Harrison WTA, Okoro CK, Asenjo JA, Andrews BA, Ferguson G, Bull AT, Goodfellow M, Ebel R, Jaspars M.** 2011. Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. J Nat Prod **74:**1491–1499. http://dx.doi.org/10.1021/np200320u.
10. **Rateb ME, Houssen WE, Harrison WTA, Deng H, Okoro CK, Asenjo JA, Andrews BA, Bull AT, Goodfellow M, Ebel R, Jaspars M.** 2011. Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment. J Nat Prod **74:**1965–1971. http://dx.doi.org/10.1021/np200470u.
11. **Nachtigall J, Kulik A, Helaly S, Bull AT, Goodfellow M, Asenjo JA, Maier A, Wiese J, Imhoff JF, Süssmuth RD, Fiedler H-P.** 2011. Atacamycins A-C, 22-membered antitumor macrolactones produced by *Streptomyces* sp. C38. J Antibiot **64:**775–780. http://dx.doi.org/10.1038/ja.2011.96.
12. **Busarakam K, Bull AT, Girard G, Labeda DP, van Wezel GP, Goodfellow M.** 2014. *Streptomyces leeuwenhoekii* sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in *Streptomyces* gene

trees. Antonie Van Leeuwenhoek **105:**849–861. http://dx.doi.org/10.1007/s10482-014-0139-y.
13. **Kang Q, Shen Y, Bai L.** 2012. Biosynthesis of 3,5-AHBA-derived natural products. Nat Prod Rep **29:**243–263. http://dx.doi.org/10.1039/C2NP00019A.
14. **Floss HG, Yu T-W, Arakawa K.** 2011. The biosynthesis of 3-amino-5-hydroxybenzoic acid (AHBA), the precursor of mC_7N units in ansamycin and mitomycin antibiotics: a review. J Antibiot **64:**35–44. http://dx.doi.org/10.1038/ja.2010.139.
15. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. **Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA.** 2000. Practical *Streptomyces* genetics. John Innes Foundation, Norwich, United Kingdom.
17. **Jones AC, Gust B, Kulik A, Heide L, Buttner MJ, Bibb MJ.** 2013. Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. PLoS One **8:**e69319. http://dx.doi.org/10.1371/journal.pone.0069319.
18. **Shima J, Hesketh A, Okamoto S, Kawamoto S, Ochi K.** 1996. Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). J Bacteriol **178:**7276–7284.
19. **Gomez-Escribano JP, Castro JF, Razmilic V, Chandra G, Andrews B, Asenjo JA, Bibb MJ.** 2015. The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. BMC Genomics **16:**485. http://dx.doi.org/10.1186/s12864-015-1652-8.
20. **Staden R, Beal KF, Bonfield JK.** 1999. The Staden package, 1998, p 115–130. *In* Misener S, Krawetz SA (ed), Bioinformatics methods and protocols. Humana Press, New York, NY.
21. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
22. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421. http://dx.doi.org/10.1186/1471-2105-10-421.
23. **Priyam A, Woodcroft BJ, Rai V, Wurm Y.** 2011. SequenceServer: BLAST search made easy. http://sequenceserver.com. Accessed 11 June 2015.
24. **Santiago-Sotelo P, Ramirez-Prado JH.** 2012. prfectBLAST: a platform-independent portable front end for the command terminal BLAST+ stand-alone suite. Biotechniques **53:**299–300. http://dx.doi.org/10.2144/000113953.
25. **Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM.** 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res **35:**W71–W74. http://dx.doi.org/10.1093/nar/gkm306.
26. **Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N.** 2013. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res **41:**D408–D414. http://dx.doi.org/10.1093/nar/gks1177.
27. **Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM, Donadio S.** 2000. Artificial chromosomes for antibiotic-producing actinomycetes. Nat Biotechnol **18:**343–345. http://dx.doi.org/10.1038/73810.
28. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
29. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
30. **Bonfield JK, Whitwham A.** 2010. Gap5—editing the billion fragment sequence assembly. Bioinformatics **26:**1699–1703. http://dx.doi.org/10.1093/bioinformatics/btq268.
31. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
32. **Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS.** 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res **33:**W455–W459. http://dx.doi.org/10.1093/nar/gki593.
33. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
34. **Bibb MJ, Findlay PR, Johnson MW.** 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene **30:**157–166. http://dx.doi.org/10.1016/0378-1119(84)90116-1.
35. **Rausch H, Lehmann M.** 1991. Structural analysis of the actinophage phi C31 attachment site. Nucleic Acids Res **19:**5187–5189. http://dx.doi.org/10.1093/nar/19.19.5187.
36. **Gregory MA, Till R, Smith MCM.** 2003. Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. J Bacteriol **185:**5320–5323. http://dx.doi.org/10.1128/JB.185.17.5320-5323.2003.
37. **Bibb MJ, White J, Ward JM, Janssen GR.** 1994. The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. Mol Microbiol **14:**533–545. http://dx.doi.org/10.1111/j.1365-2958.1994.tb02187.x.
38. **Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A.** 2011. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. Appl Environ Microbiol **77:**5370–5383. http://dx.doi.org/10.1128/AEM.00434-11.
39. **Sherwood EJ, Bibb MJ.** 2013. The antibiotic planosporicin coordinates its own production in the actinomycete *Planomonospora alba.* Proc Natl Acad Sci U S A **110:**E2500–E2509. http://dx.doi.org/10.1073/pnas.1305392110.
40. **Muth G, Nußbaumer B, Wohlleben W, Pühler A.** 1989. A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. Mol Gen Genet **219:**341–348. http://dx.doi.org/10.1007/BF00259605.
41. **Seipke RF, Loria R.** 2009. Hopanoids are not essential for growth of *Streptomyces scabies* 87-22. J Bacteriol **191:**5216–5223. http://dx.doi.org/10.1128/JB.00390-09.
42. **Poralla K, Muth G, Hartner T.** 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). FEMS Microbiol Lett **189:**93–95. http://dx.doi.org/10.1111/j.1574-6968.2000.tb09212.x.
43. **Yu TW, Müller R, Müller M, Zhang X, Draeger G, Kim CG, Leistner E, Floss HG.** 2001. Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699. J Biol Chem **276:**12546–12555. http://dx.doi.org/10.1074/jbc.M009667200.
44. **Wilson MC, Gulder TAM, Mahmud T, Moore BS.** 2010. Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the *sare1259*-encoded cytochrome P450. J Am Chem Soc **132:**12757–12765. http://dx.doi.org/10.1021/ja105891a.
45. **Yuan H, Zhao W, Zhong Y, Wang J, Qin Z, Ding X, Zhao G-P.** 2011. Two genes, *rif15* and *rif16*, of the rifamycin biosynthetic gene cluster in *Amycolatopsis mediterranei* likely encode a transketolase and a P450 monooxygenase, respectively, both essential for the conversion of rifamycin SV into B. Acta Biochim Biophys Sin (Shanghai) **43:**948–956. http://dx.doi.org/10.1093/abbs/gmr091.
46. **August PR, Tang L, Yoon YJ, Ning S, Müller R, Yu TW, Taylor M, Hoffmann D, Kim CG, Zhang X, Hutchinson CR, Floss HG.** 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. http://dx.doi.org/10.1016/S1074-5521(98)90141-7.
47. **Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T.** 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res **41:**W204–W212. http://dx.doi.org/10.1093/nar/gkt449.
48. **Minowa Y, Araki M, Kanehisa M.** 2007. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. J Mol Biol **368:**1500–1517. http://dx.doi.org/10.1016/j.jmb.2007.02.099.
49. **Yadav G, Gokhale RS, Mohanty D.** 2003. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J Mol Biol **328:**335–363. http://dx.doi.org/10.1016/S0022-2836(03)00232-8.
50. **Admiraal SJ, Walsh CT, Khosla C.** 2001. The loading module of rifamy-

cin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. Biochemistry **40:**6116–6123. http://dx.doi.org/10.1021/bi010080z.

51. **Admiraal SJ, Khosla C, Walsh CT.** 2002. The loading and initial elongation modules of rifamycin synthetase collaborate to produce mixed aryl ketide products. Biochemistry **41:**5313–5324. http://dx.doi.org/10.1021/bi0200312.
52. **Bevitt DJ, Cortes J, Haydock SF, Leadlay PF.** 1992. 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythrae*a. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. Eur J Biochem **204:**39–49.
53. **Valenzano CR, You Y-O, Garg A, Keatinge-Clay A, Khosla C, Cane DE.** 2010. Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. J Am Chem Soc **132:**14697–14699. http://dx.doi.org/10.1021/ja107344h.
54. **Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R.** 2003. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry **42:**72–79. http://dx.doi.org/10.1021/bi0268706.
55. **Zheng J, Keatinge-Clay AT.** 2011. Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. J Mol Biol **410:**105–117. http://dx.doi.org/10.1016/j.jmb.2011.04.065.
56. **Yu T-W, Shen Y, Doi-Katayama Y, Tang L, Park C, Moore BS, Hutchinson CR, Floss HG.** 1999. Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. Proc Natl Acad Sci U S A **96:**9051–9056. http://dx.doi.org/10.1073/pnas.96.16.9051.
57. **Claxton HB, Akey DL, Silver MK, Admiraal SJ, Smith JL.** 2009. Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. J Biol Chem **284:**5021–5029. http://dx.doi.org/10.1074/jbc.M808604200.
58. **Doi-Katayama Y, Yoon YJ, Choi CY, Yu TW, Floss HG, Hutchinson CR.** 2000. Thioesterases and the premature termination of polyketide chain elongation in rifamycin B biosynthesis by *Amycolatopsis mediterranei* S699. J Antibiot **53:**484–495. http://dx.doi.org/10.7164/antibiotics.53.484.
59. **Xu J, Wan E, Kim C-J, Floss HG, Mahmud T.** 2005. Identification of tailoring genes involved in the modification of the polyketide backbone of rifamycin B by *Amycolatopsis mediterranei* S699. Microbiology **151:**2515–2528. http://dx.doi.org/10.1099/mic.0.28138-0.
60. **Xiong Y, Wu X, Mahmud T.** 2005. A homologue of the *Mycobacterium tuberculosis* PapA5 protein, Rif-Orf20, is an acetyltransferase involved in the biosynthesis of antitubercular drug rifamycin B by *Amycolatopsis mediterranei* S699. Chembiochem **6:**834–837. http://dx.doi.org/10.1002/cbic.200400387.
61. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.
62. **Galperin MY.** 2010. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol **13:**150–159. http://dx.doi.org/10.1016/j.mib.2010.01.005.
63. **Absalón AE, Fernández FJ, Olivares PX, Barrios-González J, Campos C, Mejía A.** 2007. RifP; a membrane protein involved in rifamycin export in *Amycolatopsis mediterranei*. Biotechnol Lett **29:**951–958. http://dx.doi.org/10.1007/s10529-007-9340-7.
64. **Floss HG, Yu T-W.** 2005. Rifamycin—mode of action, resistance, and biosynthesis. Chem Rev **105:**621–632. http://dx.doi.org/10.1021/cr030112j.
65. **Grant SG, Jessee J, Bloom FR, Hanahan D.** 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc Natl Acad Sci U S A **87:**4645–4649. http://dx.doi.org/10.1073/pnas.87.12.4645.
66. **MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T.** 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene **111:**61–68. http://dx.doi.org/10.1016/0378-1119(92)90603-M.
67. **Gomez-Escribano JP, Bibb MJ.** 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol **4:**207–215. http://dx.doi.org/10.1111/j.1751-7915.2010.00219.x.
68. **Alting-Mees MA, Short JM.** 1989. pBluescript II: gene mapping vectors. Nucleic Acids Res **17:**9494. http://dx.doi.org/10.1093/nar/17.22.9494.
69. **Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE.** 1992. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene **116:**43–49. http://dx.doi.org/10.1016/0378-1119(92)90627-2.
70. **Rodríguez-García A, Santamarta I, Pérez-Redondo R, Martín JF, Liras P.** 2006. Characterization of a two-gene operon *epeRA* involved in multidrug resistance in *Streptomyces clavuligerus*. Res Microbiol **157:**559–568. http://dx.doi.org/10.1016/j.resmic.2005.12.008.
71. **Hong H-J, Hutchings MI, Hill LM, Buttner MJ.** 2005. The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. J Biol Chem **280:**13055–13061. http://dx.doi.org/10.1074/jbc.M413801200.
72. **Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH.** 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res **43:**D222–D226. http://dx.doi.org/10.1093/nar/gku1221.

Appendix K

Research article *in extenso*: "The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2"

J. P. Gomez-Escribano, **J. F. Castro**, V. Razmilic, G. Chandra, B. Andrews, J. A. Asenjo and M. J. Bibb.
BMC Genomics, 2015, volume 16, issue 1, page 485.
DOI 10.1186/s12864-015-1652-8
Supplementary information available at BMC website:
http://www.biomedcentral.com/1471-2164/16/485/additional.

Gomez-Escribano et al. BMC Genomics (2015) 16:485
DOI 10.1186/s12864-015-1652-8

BMC Genomics

RESEARCH ARTICLE Open Access

The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2

Juan Pablo Gomez-Escribano[1*], Jean Franco Castro[1,2], Valeria Razmilic[1,2], Govind Chandra[1], Barbara Andrews[2], Juan A. Asenjo[2] and Mervyn J. Bibb[1]

Abstract

Background: Next Generation DNA Sequencing (NGS) and genome mining of actinomycetes and other microorganisms is currently one of the most promising strategies for the discovery of novel bioactive natural products, potentially revealing novel chemistry and enzymology involved in their biosynthesis. This approach also allows rapid insights into the biosynthetic potential of microorganisms isolated from unexploited habitats and ecosystems, which in many cases may prove difficult to culture and manipulate in the laboratory. *Streptomyces leeuwenhoekii* (formerly *Streptomyces* sp. strain C34) was isolated from the hyper-arid high-altitude Atacama Desert in Chile and shown to produce novel polyketide antibiotics.

Results: Here we present the *de novo* sequencing of the *S. leeuwenhoekii* linear chromosome (8 Mb) and two extrachromosomal replicons, the circular pSLE1 (86 kb) and the linear pSLE2 (132 kb), all in single contigs, obtained by combining Pacific Biosciences SMRT (PacBio) and Illumina MiSeq technologies. We identified the biosynthetic gene clusters for chaxamycin, chaxalactin, hygromycin A and desferrioxamine E, metabolites all previously shown to be produced by this strain (*J Nat Prod*, 2011, 74:1965) and an additional 31 putative gene clusters for specialised metabolites. As well as gene clusters for polyketides and non-ribosomal peptides, we also identified three gene clusters encoding novel lasso-peptides.

Conclusions: The *S. leeuwenhoekii* genome contains 35 gene clusters apparently encoding the biosynthesis of specialised metabolites, most of them completely novel and uncharacterised. This project has served to evaluate the current state of NGS for efficient and effective genome mining of high GC actinomycetes. The PacBio technology now permits the assembly of actinomycete replicons into single contigs with >99 % accuracy. The assembled Illumina sequence permitted not only the correction of omissions found in GC homopolymers in the PacBio assembly (exacerbated by the high GC content of actinomycete DNA) but it also allowed us to obtain the sequences of the termini of the chromosome and of a linear plasmid that were not assembled by PacBio. We propose an experimental pipeline that uses the Illumina assembled contigs, in addition to just the reads, to complement the current limitations of the PacBio sequencing technology and assembly software.

Keywords: Second/Third next generation sequencing, Illumina MiSeq, Pacific Biosciences PacBio SMRT, Chaxamycin, Chaxalactin, Lasso peptide, Genome mining

* Correspondence: Juan-Pablo.Gomez-Escribano@jic.ac.uk
[1]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom
Full list of author information is available at the end of the article

Background

Actinomycetes are Gram-positive mycelial bacteria found predominantly in soil, but they also occur in symbiotic associations with terrestrial and aquatic invertebrates. They undergo a complex process of morphological and physiological differentiation that leads to the production of exospores and specialised metabolites [1] with a wide range of biological activities. While the function of many of these molecules in the natural environment is not always evident, they are believed to provide a competitive advantage to the producing organism [2]. Many of these specialised metabolites possess potent antibiotic activity, and actinomycetes produce over 70 % of the natural product scaffolds found in clinically used anti-infective agents [3].

The pioneering sequencing and analysis of the genomes of *Streptomyces coelicolor* A3(2) [4] and *Streptomyces avermitilis* [5] revealed that actinomycetes possess an unexpected abundance of natural product biosynthetic gene clusters and thus that they have the potential to make many more compounds than previously thought. This led to the "genome mining" approach to natural product discovery, where bioinformatic analysis is used to estimate the biosynthetic capacity, and potential metabolite novelty, of an organism before extensive analysis in the laboratory. Genome mining of actinomycetes and other microorganisms has already provided access to many novel biosynthetic pathways and metabolites that otherwise would have remained undetected [6, 7]. It may be particularly useful when analysing novel and possibly difficult to culture microorganisms isolated from unusual and unexploited habitats such as the oceans, deserts, and the surfaces of plants and animals.

To be carried out efficiently, genome mining relies on the availability of a good quality genome sequence obtained at an affordable price and in a short time-frame. Until recently, whole genome shotgun sequencing using Next-Generation Sequencing (NGS) technologies could not be expected to yield less than several hundred contigs for an actinomycete genome of 6 – 12 Mb, with biosynthetic gene clusters often split over several contigs. Much bioinformatic analysis and contig-stitching by, for example, PCR would often be required to identify a complete cluster. Moreover, the short read length of Second Generation Sequencing technologies (like Illumina) makes it very difficult to correctly assemble the long repetitive sequences typically found in biosynthetic gene clusters containing modular polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs). In such clusters, it is not unusual to find regions of high homology between genes and intragenic tandem repeats of 650-1000 bp with close to 100 % nucleotide sequence identity, well beyond the read length provided by Illumina (up to 2 × 300 bases); e.g., the *S. coelicolor* coelimycin PKS gene *sco6274*, positions 3879-4533 and 11986- 12639; and the calcium-dependent antibiotic NRPS gene *sco3230*, positions 13187-14121 and 16307-17241. The long sequence reads provided by Third Generation Sequencing techniques such as the Pacific Biosciences SMRT technology should allow more reliable assemblies of PKS and NRPS gene clusters and, in principle, yield single contigs for all of the replicons present in an actinomycete. The SMRT technology also produces more even sequence coverage of the high mol% G + C DNA found in actinomycete genomes than other NGS platforms [8].

Streptomyces leeuwenhoekii (formerly *Streptomyces* sp. C34) was isolated, together with many other novel actinomycetes, from the saline Chaxa Lagoon in the high-altitude Atacama Desert in northern Chile [9, 10]. It produces previously described metabolites (the siderophore desferioxamine E and the antibiotic hygromycin A), but also novel polyketide antibiotics, the chaxamycins and chaxalactins. Chaxamycin A – D are four novel ansamycin-type polyketides with promising antibacterial activity against MRSA and anti-proliferative activity resulting from the inhibition of the ATPase activity of the human Hsp90 protein [11, 12]. A draft genome sequence of this strain with 658 contigs had been derived previously from Illumina GA IIx 100 bp paired-end reads [10]; attempts to mine this sequence for the chaxamycin and chaxalactin biosynthetic gene clusters (Castro *et al.*, submitted) revealed many inconsistencies and misassemblies (see Additional file 1), presumably a consequence of the short read lengths of the Illumina technology and the repetitive nature of the two gene clusters.

Given the unusual origin of *S. leeuwenhoekii* and our desire to analyse its biosynthetic potential, we set out to sequence its genome using the most advanced technology available. This has allowed us to generate an almost complete chromosome sequence as well as the sequences of two plasmids as single contigs without recourse to gap-closing or sequencing of clones from a genomic library; to our knowledge, this is the first time that this has been achieved with an actinomycete. We also report here our findings on the advantages and limitations of the two technologies we used, Illumina MiSeq and Pacific Biosciences RSII, and we propose a pipeline for the generation of high quality actinomycete genome sequences.

Results and Discussion

Availability of data

The fully annotated sequences presented in this work have been deposited in the European Nucleotide Archive under Study accession number PRJEB8583 (http://www.ebi.ac.uk/ena/data/view/PRJEB8583). The sequences of pSLE1, pSLE2 and the chromosome have been assigned accession numbers LN831788, LN831789 and LN831790 respectively.

Sequencing and assembly of the *S. leeuwenhoekii* genome

Two technologies were used to sequence genomic DNA isolated from *S. leeuwenhoekii*: Illumina MiSeq (as available in August 2013) and Pacific Biosciences (PacBio) RSII (as available in November 2013). Assembly of the Illumina sequencing data yielded 279 contigs, assembled into 175 scaffolds, totalling 8064420 bp. Assembly of the PacBio sequencing data produced three contigs of 7895833, 9613 and 94746 bp, totalling 8000192 bp. The PacBio contig of 7.9 Mb was expected to contain most of the sequence of the chromosome, and was referred to as C34-chromosome version 1. The small 9613 bp contig from the PacBio assembly was found to match the 7.9 Mb contig from position 5938858 to 5929122 (reverse complement) but with only 92 % identity, possibly indicating that it originated from reads with accumulated errors. These small, error-prone, contigs have been observed by others using PacBio sequencing (Silke Alt, Natalia Miguel-Vior, Zhiwei Qin and Thomas Scott, personal communications). This contig was discarded from any further analysis. A detailed description of the sequencing and bioinformatic analysis of this and following sections can be found in Additional file 2 – Materials and Methods.

Correction of the PacBio 7.9 Mb contig (C34-chromosome version 1) using the Illumina assembly

Comparison of the Illumina MiSeq and PacBio assemblies readily identified problems of misassembly in the Illumina contigs, and in particular in the regions containing polyketide biosynthetic gene clusters, a particular focus of our initial interest in *S. leeuwenhoekii* (see Additional file 1: Figure S3). However, aware of the higher nucleotide-accuracy of the Illumina technology compared to PacBio, we mapped the Illumina MiSeq contigs (not scaffolds) over the PacBio-generated C34-chromosome version 1. The generated alignment was manually edited with GAP5 [13] to correct the PacBio sequence with the Illumina contigs while accommodating possible misassembles and systematic errors inherent in the Illumina technology [14]. Most of the differences were apparent omissions in the PacBio sequence of a C or G in homopolymeric runs of three or more Cs or Gs. These omissions were confirmed by analysis of individual sequence differences; the missing nucleotides in the PacBio sequence resulted in frame shifts that were readily identified using GC-Frame Plot [15] in Artemis [16] and substantiated by inspection of the corresponding amino acid sequences. Our analysis indicated that the Illumina sequence always contained the correct number of bases in these homopolymeric stretches, and consequently it was used to correct the PacBio assembly. This resulted in the insertion of 2934 bases and an additional 42 base changes, a total of 2976 corrections in a final chromosomal assembly of 7898767 bases (0.03768 %). This sequence was referred to as C34-chromosome version 2.

Extension of 7.9 Mb corrected-contig (C34-chromosome version 2) with the Illumina assembly

During the previous correction step we identified sequence in the Illumina assembly that was absent from the PacBio assembly. This was achieved by using all of the Illumina assembled data instead of the Illumina reads, which is the usual bioinformatic practice (in e.g., software like iCORN [17] and Mira [18]). We found that Illumina contig 0089 (29048 bp) mapped at the 5'-end of the PacBio 7.9 Mb contig but contained 5.1 kb of additional sequence (see Additional file 3: Figure S3). After careful examination of the genetic content of contig 0089, we concluded that the extra 5.1 kb was genuine *S. leeuwenhoekii* sequence likely located at the end of the terminal inverted repeat (see below). BLAST analysis [19] and alignment in GAP5 were then used to extend the 5'-end of the chromosomal sequence with an extra 5121 bases to yield a chromosome of 7903888 bp, referred to as C34-chromosome version 3.

Further correction of C34-chromosome version 3 with the Illumina paired-end reads

The original, unassembled, Illumina paired-end reads with quality values were aligned to version 3 of the chromosome with two different programs, BWA [20, 21] and Bowtie 2 [22], processed with SAMtools and bcftools [23] to call the potential variants, and these were then studied within GAP5. We focused on the additions/omissions reported by both alignment programs. While we decided not to incorporate any of the reported base changes, we did correct a few additions/omissions, in particular an erroneous addition at position 7169599 that was only reported by the BWA alignment (GTGGA was corrected to GTGA) which repaired a reading-frame shift in a polyketide synthase gene that is part of the chaxalactin biosynthetic gene cluster (Castro *et al.*, manuscript in preparation). Three corrections were also made at the beginning of the proposed Terminal Inverted Repeat (TIR) at the 5'-end of the sequence; a C was added in a run of five Cs after the G at position 387475, giving six Cs; a C was added after A at new position 387818 (acccaa to aCcccaa); and a C was added after A at new position 387824 (aacccca to aaCcccca). Finally, a G was added at position 7383828 (cggg to cGggg), correcting an omission reported by both programs that introduced a frame shift in an uncharacterised polyketide synthase gene (cluster 28 of Table 1).

The final chromosomal sequence, referred to as C34-chromosome version 4, contains 7903895 nucleotides and has a mol% G + C content of 72.76 %.

Annotation of the *Streptomyces leeuwenhoekii* chromosome

Annotation of the *S. leeuwenhoekii* chromosome version 4 was performed using Prodigal [24] to identify protein

Table 1 Putative biosynthetic gene clusters for specialised metabolites

antiSMASH Cluster No.	antiSMASH type descriptor	Position (*manually annotated clusters in italics*)		Our annotation (based on Ref.)
		From	To	
1	T1pks	99264	143430	
	Not identified	*160425*	*189028*	Hygromycin A [11, 44]
2	T1pks	191701	240196	
3	T1pks-nrps	324784	392261	
4	Nrps	379508	426758	
5	T3pks	416888	458084	
6	Bacteriocin	572464	582679	
7	Terpene	598795	619823	
	Not identified	*684373*	*654830*	Lasso-peptide 2
8	Nrps	714060	794426	
9	Terpene	1056004	1076960	
10	T2pks-transatpks-nrps	1075399	1155931	Halogenated polyketide [Razmilic et al.]
11	*T1pks-terpene*	*1211049*	*1289829*	Chaxamycin [Castro et al.]
12	T1pks	1497127	1544539	
13	Terpene	1624097	1645110	
14	T1pks-siderophore	1776281	1833813	
15	Terpene	1972277	1994487	
16	Bacteriocin	2013690	2025087	
17	Siderophore	2293580	2305424	Highly conserved
18	Nrps-t1pks	2668194	2719415	
19	T3pks	2937137	2978264	
20	*Terpene*	*3056325*	*3058819*	Albaflavenone [45, 46]
	Not identified	3560196	3564842	Lasso-peptide 1
21	*Siderophore*	*5237176*	*5244356*	Desferrioxamine E [11, 47, 48]
22	Melanin	5330379	5340933	
23	Amglyccycl-butyrolactone	5385171	5417416	
24	Ectoine	6176293	6186691	
25	Other	6710095	6751819	
26	T3pks	6822979	6864043	
27	T1pks	7141058	7240871	Chaxalactin [11, Castro et al.]
28	T1pks	7355977	7439461	
29	Other	7486047	7529121	
30	Terpene-t2pks	7530162	7588405	
31	Terpene	7744176	7768730	
	Not identified	*pSLE2 103389*	*pSLE2 105999*	Lasso-peptide 3

coding sequences (PCS) followed by BASys [25] for assignment of putative function. We chose gene identifiers starting with "sle", for "*Streptomyces leeuwenhoekii*", followed by five digits starting with "sle_00010" with increments of 10 to allow for the addition of subsequently identified genes and genetic features. Neither Prodigal nor BASys annotated rRNA or tRNA genes, so the chromosome sequence was submitted to the RAST server [26, 27] and the rRNA and tRNA annotations added to generate the published *S. leeuwenhoekii* chromosome sequence.

Identification of the possible Terminal Inverted Repeats (TIR) and chromosome ends

Streptomyces species frequently possess linear chromosomes with Terminal Inverted Repeats (TIRs) of almost identical sequence with covalently-bound terminal proteins for priming of replication [28]. The length of the

TIR is very variable among species, ranging from only 14 bp in *Streptomyces hygroscopicus* 5008 [29] to over 1 Mbp in *S. coelicolor* [30].

BLAST analysis and examination in Artemis revealed that the last 6996 bases at the 3'-end of the chromosome sequence (hereinafter referred to as "right TIR") were 99 % identical to a segment of reverse-complementary sequence around position 388 kb (hereinafter referred to as "left TIR"; see Additional file 3, Fig. 1). It is likely that this 7 kb sequence at the right end represents the start of the right TIR, and was acquired by some of the long PacBio reads that extended from the region next to the TIR, while the rest of the reads from the right TIR have probably been assembled with the left TIR. Visualisation of the PacBio assembly BAM file revealed a pronounced increase in coverage of the first 380 kb of the assembly (see Additional file 3: Figure S2), probably reflecting the incorporation of reads from both the left and right TIR in the assembly of the left TIR.

Detailed analysis of the end of the left TIR (the start of C34-chromosome version 4) identified two genes, *sle_00020* and *sle_00120*, with 84 % nucleotide sequence identity to each other, and predicted amino acid sequences that were 73 % and 72 % identical, respectively, to a helicase-like protein encoded by *sco0002* of *S. coelicolor*, homologues of which are present and highly conserved at both ends of all sequenced linear replicons (chromosomes and plasmids) from *Streptomyces* species [31, 32]. The presence of two contiguous helicase-like encoding genes has not been described before but detailed sequence inspection failed to reveal possible misassembly. Mfold [33] analysis of the DNA sequence upstream of *sle_00020* (but not *sle_00120*) revealed palindromic repeats with the potential to form a complex secondary structure (see Additional file 3: Figure S3) similar to those found in other *Streptomyces* linear replicons [34].

Our analysis suggests that the TIRs of the *S. leeuwenhoekii* chromosome extend for about 388 kb, at the upper end of published *Streptomyces* TIR sequences. This also implies that the hygromycin A biosynthetic gene cluster, located between position 160-189 kb, could be duplicated at each end of the chromosome should both TIRs show complete gene conservation. While still infrequent, biosynthetic gene clusters for specialised metabolites localised in the TIRs of actinomycetes have been described previously [35, 36] and demonstrated experimentally [35].

Assembly and annotation of pSLE1, an 86.4 kb circular plasmid

The 94746 bp PacBio-contig contained direct repeats of 8.4 kb present at both ends, indicative of an 86 kb circular DNA molecule. Most of this sequence was also found in a contig of 86370 bp in the Illumina assembly. BLAST analysis using the non-redundant NCBI database suggested that it represents a plasmid, which we called pSLE1.

A detailed description of the assembly of the circular DNA sequence using both PacBio and Illumina data is given in the Additional file 2 and Additional file 4. pSLE1 possesses a base composition of 69 % mol% G + C. 133 putative PCSs were identified with Prodigal and tagged as "sle2_001"; putative functions were added manually in Artemis using BLAST and Pfam [37] searches.

For most of the PCSs, in particular those encoding putative phage particle proteins, the annotation was based on that of pZL12 [38], a well characterised plasmid with high levels of homology and synteny to pSLE1. Genes encoding ParA and ParB partitioning proteins were identified with Pfam, supporting the proposition that this contig represents an autonomously replicating element. pSLE1 ParA shows no discernible identity to the chromosomal ParA, but is 62 % identical to pZL12.17c ParA and 67 % identical to the putative plasmid partitioning protein ParA from *Streptomyces rochei* (sequence ID BAK19858). pSLE1 ParB shares 31 % identity with the chromosomal ParB, but 41 % identity with pZL12.16c ParB and 67 % identity with the putative plasmid partitioning protein ParB from *S. rochei* (sequence ID BAK19859). Interestingly, pSLE1 contains two genes encoding putative integrases, suggesting that the plasmid might be capable of insertion into the chromosome.

A search with antiSMASH [39] did not identify any putative specialised metabolite gene clusters in pSLE1 (a gene encoding a putative Type III PKS with 59 % amino acid sequence identity to pZL12-100 and similar to RppA from *Streptomyces antibioticus* was identified but, as in pZL12, the neighbouring genes did not indicate that it was part of a natural product gene cluster).

Assembly and annotation of pSLE2, a >132 kb linear plasmid

Two Illumina contigs, contig0026 with 97005 bp and contig0079 with 32697 bp, did not match any sequence in the PacBio assembly. BLAST searches against the non-redundant NCBI database indicated that it could represent a plasmid, which we designated pSLE2. We were subsequently able to join the Illumina contigs using corrected PacBio reads to yield a 132 kb sequence (see Additional file 2 and Additional file 5 for details).

pSLE2 has a base composition of 70 mol% G + C. PCSs were identified using Prodigal and tagged as "sle2_001"; putative functions were manually annotated in Artemis using BLAST and Pfam searches.

We could not identify a clear direct repeat indicative of a circular plasmid, as we did for pSLE1, or a putative TIR characteristic of a linear plasmid, suggesting that some sequence might be missing. However, the 5'-end of

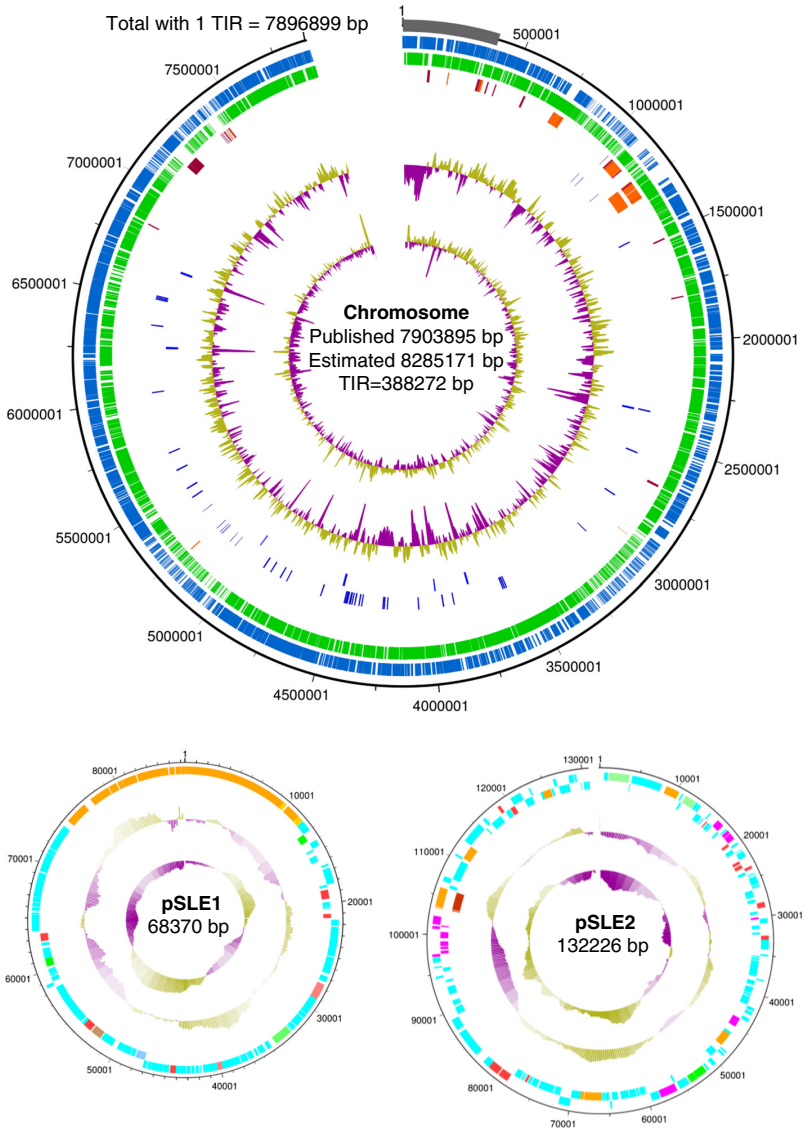


Fig. 1 Schematic representation of the *S. leeuwenhoekii* chromosome, circular plasmid pSLE1, and linear plasmid pSLE2 (incomplete sequence). The chromosome is represented as an open circle, covering only the published sequence without the duplication of the terminal inverted repeat (represented as a grey band starting at position 1). From outside to inside, the concentric circles represent: nucleotide position; Protein Coding Sequences (PCSs) on the forward strand; PCSs on the reverse strand; PCSs for putative biosynthetic genes for specialised metabolites (dark red indicates the forward strand, orange the reverse strand); the orange box shown in the fifth circle indicates the chaxamycin biosynthetic gene cluster; tRNA and rRNA genes are shown in the sixth and seventh lines, respectively, in dark blue; the eighth concentric circle shows the GC-plot (GC %, window size = 10000; base step size = 200) and the inner-most circle the GC-skew ([(G – C)/(G + C)] window size = 10000; base step size = 200), both calculated using the sequence with both TIRs, a window size of 10000 and a step size of 200 (purple and olive indicate below and above average, respectively). For pSLE1 and pSLE2, PCSs are coloured red for putative regulatory genes; green, for plasmid replication and partitioning genes; the fourth circle in shows the GC-plot and the inner-most circle the GC-skew, both calculated as for the chromosome. For pSLE1, phage-related genes are shown in orange, and the type III PKS (chalcone synthase) gene is shown in brown. For pSLE2, genes with known plasmid functions are in orange; genes annotated as mobile elements and involved in transposition are in pink; the lasso-peptide biosynthetic gene cluster is shown in dark orange. Not to scale

pSLE2 (provided by the 5'-end of Illumina contig 0026) is highly similar (over 85 % nucleotide identity) to the end of the TIR identified for the chromosome (see Additional file 5: Figure S2), and includes a gene for a putative terminal helicase (*sle2_002*) with 87 % and 85 % amino acid identity to Sle_00020 and Sle_00120, respectively. Such high identity between linear chromosomes and linear plasmids is a feature of *Streptomyces* species [31] and suggests that pSLE2 is a linear plasmid and that this sequence is at one end of a TIR.

Two contiguous genes were found by Pfam searches to encode possible functional homologues of the partition proteins ParA and ParB (no significant identity was found to any of the putative partitioning proteins encoded by the chromosome or pSLE1, further suggesting that this is a separate replicon). Pulse-Field Gel-Electrophoresis (PFGE) of total DNA isolated from *S. leeuwenhoekii* revealed an extrachromosomal replicon with an apparent linear size of between 112 and 130.5 kb (Additional file 5: Figure S1), consistent with the predicted size of pSLE2.

As for pSLE1, we found a gene, *sle2_153*, encoding a putative integrase, suggesting that the plasmid might be capable of integrating into the chromosome. In addition, we identified at least three genes encoding possible conjugation functions (*sle2_062*, *sle2_090* and *sle2_091*) suggesting that the plasmid might be capable of self-transmission by conjugation.

An interesting feature is the presence of a gene, *sle2_140*, encoding a putative zeta-toxin (Pfam family "Zeta_toxin" (PF06414) e-value 1.9e-45) potentially involved in ensuring the maintenance and segregation of the plasmid [40]; we did not find a gene encoding a putative anti-toxin, but we did find a toxin-antitoxin system encoded in the chromosome similar to that reported for *S. coelicolor* and *Streptomyces lividans* [41].

AntiSMASH did not reveal any putative natural product gene clusters, but we did find a gene cluster potentially encoding the biosynthesis of a lasso-peptide, which we have subsequently identified and characterised (manuscript in preparation).

Intriguingly, there are several genes encoding putative transposases, many with frame shifts (revealed using GC Frame Plot). This region was not present in the Illumina assembly, and lies between Illumina contigs 0026 and 0079; only the PacBio corrected-reads contained the full sequence (see Additional file 2). The presence of many repeated sequences with high levels of nucleotide sequence identity may explain their absence from the Illumina assembly.

General characteristics of the genome sequence

Our assembled chromosome sequence contains 7903895 bp, with a mol% G + C content of 73 %, consistent with other members of the genus *Streptomyces*. With the proposed addition of the right hand 388 kb TIR, the predicted size of the genome would be 8285171 bp (our 7903895 bp assembly, minus 6996 bp of the start of right TIR, plus the duplication of the left TIR of 388272 bp, equals 8285171 bp for the chromosome with two equal 388 kb TIRs). This is similar to the size of many streptomycete genomes (e.g., the chromosome of *S. coelicolor* is 8667507 bp [1]).

6712 PCSs were predicted for our assembled sequence, and 7057 PCSs if we include the PCSs predicted for the right TIR. A table of COG (Clusters of Orthologous Genes) functional categories [42] was determined (Table 2).

The *S. leeuwenhoekii* chromosome contains six rRNA operons located between nucleotide positions 2364598 and 6687647; interestingly, an orphan gene encoding a 5S rRNA was found at position 3962518. DnaA and DnaN replication proteins are encoded in genes spanning nucleotide positions 4260071-4258170 and 4256851-4255721, respectively, indicating that the likely origin of replication of the chromosome lies between 4256851 and 4258170. Consistent with precedent, the two rRNA operons that lie between nucleotide positions 2364598 and 3779582 (and the 5S rRNA orphan gene at position 3962518) are located on the complementary strand, while the four rRNA operons located between nucleotide positions 4473032 and 6687647 are located on the top strand (see Additional file 6: Table S1). We identified 65 tRNA genes (see Additional file 6: Table S2), similar to the number found in other *Streptomyces* (e.g., the *S. coelicolor* genome contains 64 tRNA genes and one tRNA pseudogene).

The general characteristics of the *S. leeuwenhoekii* genome are summarised in Table 3.

Biosynthetic gene clusters for specialised metabolites

Putative gene clusters encoding the biosynthesis of specialised metabolites were identified with antiSMASH 2 [39] and non-automated searches with BLAST and Pfam (Table 1). *S. leeuwenhoekii* was known to produce the siderophore desferrioxamine E and the antibiotic hygromycin A [11]; we identified probable biosynthetic gene clusters for both of these metabolites by searching for homologues of the corresponding gene clusters from *S. coelicolor* and *Streptomyces hygroscopicus* NRRL 2388, respectively. The desferrioxamine E gene cluster extends from *sle_44550* to *sle_44600* and shows complete gene synteny with the homologous gene cluster from *S. coelicolor* with protein identities varying between 77 % and 93 %, while the putative hygromycin A gene cluster, which spans from *sle_01610* to *sle_01870*, shows almost complete synteny with the homologous gene cluster from *S. hygroscopicus*, with protein identities of 64 % to 90 %. Interestingly, antiSMASH did not identify the hygromycin A gene cluster which lies within the proposed left TIR; two putative type-I PKS gene clusters

Table 2 COG functional categories. COG (Clusters of Orthologous Genes) functional categories of chromosomal protein codding sequences identified in *S. leeuwenhoekii* chromosome, and from *S. coelicolor* for comparison (as calculated by BASys [25] for both genomes)

	Streptomyces leeuwenhoekii		*Streptomyces coelicolor*	
COG functional categories	Percentage	Number	Percentage	Number
Energy production and conversion	4	270	4.1	317
Cell division and chromosome partitioning	0.5	30	0.4	31
Amino acid transport and metabolism	6	400	5.1	395
Nucleotide transport and metabolism	1.3	89	1.2	93
Carbohydrate transport and metabolism	6.5	433	6.5	503
Coenzyme metabolism	2.3	154	2.2	170
Lipid metabolism	3.7	250	3.3	255
Translation, ribosomal structure and biogenesis	2.7	182	2.5	193
Transcription	7.3	493	8.5	658
DNA replication, recombination and repair	2.9	193	2.9	224
Cell envelope biogenesis, outer membrane	3	204	2.9	224
Cell motility	0.1	5	0.1	8
Posttranslational modification, protein turnover, chaperones	1.9	127	1.8	139
Inorganic ion transport and metabolism	2.3	154	2.7	209
Secondary Structure	2.8	186	1.9	147
General function prediction only	6.9	460	7.4	572
COG of unknown function	3.6	243	3.6	278
Signal Transduction	3.8	256	4.2	325
Unknown	36.6	2458	36.7	2839

and one putative hybrid type-I PKS-NRPS gene cluster were also found in the left TIR. If, as we have proposed, the TIRs span for 388 kb, then all four of these gene clusters would be duplicated at the other end of the chromosome.

S. leeuwenhoekii is also known to produce two novel families of polyketide antibiotics, the chaxalactins and the ansamycin-like chaxamycins [11, 12]. The biosynthetic gene clusters for both antibiotics were identified and are currently under investigation.

Table 3 General characteristics of the *S. leeuwenhoekii* genome

Assembled chromosome size	7903895 bp
Estimated chromosome size	8285171 bp
Estimated Terminal Inverted Repeats	388272 bp
Chromosome topology	Linear
Chromosome G + C content	73 %
rRNA operons	6
tRNA genes	65
pSLE1 circular plasmid	86370 bp
pSLE1 G + C content	69 %
pSLE2 linear plasmid	132226 bp
pSLE2 G + C content	70 %
Putative biosynthetic gene clusters for specialised metabolites	34 (+1 in pSLE2)

We also identified three putative gene clusters each encoding the biosynthesis of unknown lasso-peptides (LS). Two of them, the "lasso-peptide 1" and "lasso-peptide 2" gene clusters (named *ls1* and *ls2*, respectively) are located on the chromosome, while the "lasso-peptide 3" gene cluster (named *ls3*) is located on the linear plasmid pSLE2. All three gene clusters are currently under study in our laboratory, where PCR amplification and Sanger sequencing have confirmed that the assembly and sequence reported here is correct. These clusters were not identified by antiSMASH.

Conclusions

Benefits and challenges of sequencing technologies: a revised pipeline

To our knowledge, this is the first report of the use of NGS to produce a high quality and non-fragmented genome sequence of an actinomycete, an essential prerequisite for efficient genome mining for natural product discovery in these GC-rich bacteria. Our assembly yielded the sequences of a single chromosomal contig, a complete circular plasmid (pSLE1), and most of a linear plasmid (pSLE2). Although the Illumina MiSeq assembly

produced a single contig for pSLE1, it would not have been possible to confirm its circularity and completeness without the PacBio data. Overall, we used two SMRT cells (plus a small amount of data from a faulty run) resulting in over 120x coverage. Thus, with the improving capacity of the PacBio SMRT cells, we predict that a maximum of three SMRT cells should be sufficient to obtain the entire complement of replicons of an actinomycete genome of around 8 Mb in single contigs. This contrasts with the previously published Illumina draft genome sequence of *S. leeuwenhoekii* [10] that contained 658 contigs in which we found many misassemblies and missing sequence (see Additional file 1).

The PacBio long reads also permitted the identification of the start of the right hand TIR, but they did not provide any sequence information for the 5 kb at the ends of the chromosome and the linear plasmid pSLE2, which was obtained from the Illumina data. The reason for this is unknown, but it could reflect the different strategies used for library construction; while PacBio focuses on large fragments of ~20 kb, the average size of the Illumina library used for sequencing was 550 bp. Also, no specific procedure was used during DNA purification to remove the protein that is covalently bound to the termini of the linear replicons, perhaps resulting in lower sequence coverage of the terminal regions that might be further exacerbated by the large fragment size used for PacBio sequencing.

Surprisingly, the PacBio assembly lacked the linear pSLE2 sequence, although the sequence information was present in the corrected PacBio reads. The Illumina assembly contained two large contigs covering most of pSLE2 that were merged into a contiguous sequence using the PacBio corrected reads, but this plasmid would have not been identified with the PacBio assembly alone.

These two findings highlight the importance of the current need to use both PacBio and Illumina assemblies, instead of assembling only the PacBio data and then using the Illumina reads to correct the assembly, which appears to be the accepted practice [17, 18, 43]. We cannot explain the lack of assembly of some of the PacBio data present in the corrected reads (it might be due to a difference in relative abundance compared to the rest of the sequence, but this will require analysis of the PacBio assembly algorithm) but this was compensated for by using the Illumina MiSeq assembly.

In summary, for a *de novo* shotgun genome sequence from an actinomycete aimed at yielding single contigs per replicon, we currently propose a strategy (Fig. 2) that includes sequencing genomic DNA with PacBio RSII using initially two (and a third later if required) SMRT cells and a >20 kb insert library (aiming at >100x coverage) combined with Illumina MiSeq paired-end sequencing of a 500 bp library without PCR amplification (to avoid introducing bias from uneven amplification of high G + C actinomycete DNA (aiming at >90x coverage)). Both data sets are assembled and compared, and the Illumina contigs used to correct the PacBio nucleotide omissions/additions, which should be confirmed using GC Frame Plot and BLAST analyses. This consensus is further corrected with the Illumina reads. Despite the highly efficient current assembly algorithms, a

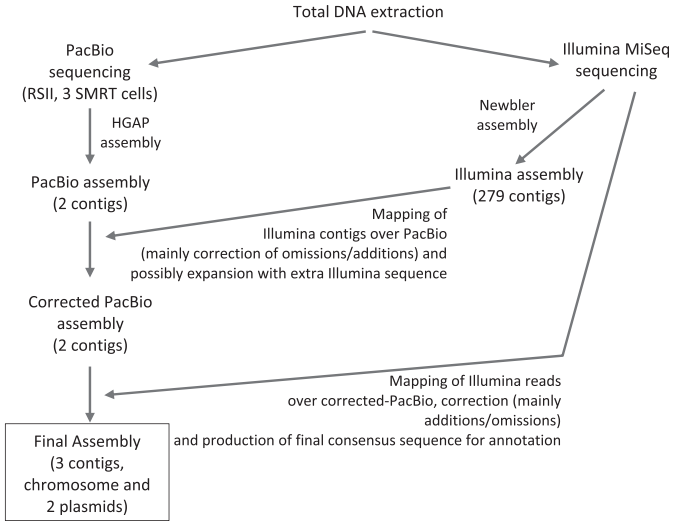


Fig. 2 Sequencing and assembly pipeline. The sequencing and assembly pipeline followed in this work (data specific to this project are shown in brackets) and suggested as strategy for actinomycete genome sequencing

considerable amount of human input was still needed to obtain a high quality single contig assembly, and accurate annotation of gene function.

Methods

Extensive details of the methodology and materials used during this study are given in Additional file 2. Perl scripts are given in Additional file 7 and Additional file 8.

Additional files

Additional file 1: Comparison of assemblies and examples of misassemblies (using the Artemis Comparison Tool). Three figures illustrating the problems of assembly of polyketide synthase genes with Illumina data.

Additional file 2: Materials and Methods. This Microsoft Word file contains all methodology and materials used to extract and sequence the genomic DNA, and the software used, pipelines, strategies and decisions followed during bioinformatic analysis. A detailed description of the process followed to compare assemblies, map reads and correct the chromosome is included, as well as a detailed description of the process used to assemble the two plasmids. It also includes details of the process used for annotation of gene function.

Additional file 3: Determination of Terminal Inverted Repeat. Three figures illustrating and supporting the identification of the Terminal Inverted Repeat.

Additional file 4: Assembly of circular plasmid pSLE1. Figure illustrating the organisation of Illumina and PacBio data covering pSLE1.

Additional file 5: Assembly of linear plasmid pSLE2. Two figures showing an extrachromosomal replicon, possibly pSLE2, after Pulse Field Gel Electrophoresis, and the similarity between the chromosome and pSLE2 left ends.

Additional file 6: rRNA operons and tRNA genes. Two tables with the identified rRNA operons and tRNA genes.

Additional file 7: Perl script "fastaToIndividualFiles.pl". Perl script (text file) used to extract from a multi-fasta file, and rename at the same time, specific sequences whose identifiers are listed in a text file (if a list is not provided, all sequences are output to individual files).

Additional file 8: Perl script "renameSequencesInFasta.pl". Perl script (text file) used to extract from a multi-fasta file, and rename including the sequence length at the same time, specific sequences whose identifiers are listed in a text file (if a list is not provided, all of the sequences are output to individual files).

Competing interests

The authors declare that they have no competing interests

Authors' contributions

JPG extracted and quality-controlled genomic DNA, analysed data and corrected the chromosome assembly, assembled and corrected both plasmids, annotated and submitted the sequences and wrote the manuscript. JFC extracted and quality-controlled genomic DNA, annotated the chromosome and contributed to writing the manuscript. VR annotated the chromosome and contributed to writing the manuscript. GC designed bioinformatic analysis, wrote the Perl scripts, prepared and maintained BLAST databases and server, and contributed to writing the manuscript. BA contributed to experimental design and to writing the manuscript. JAA contributed to experimental design and to writing the manuscript. MJB designed the overall strategy, analysed data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Michael Goodfellow and Alan Bull for providing *S. leeuwenhoekii*, and to Giles van Wezel and Geneviève Girard from the University of Leiden for providing access to the draft sequence of *S. leeuwenhoekii* prior to publication. J. F. C. and V. R. received National PhD Scholarships (#21110356 and #21110384, respectively) and Visiting Student Scholarships (Becas Chile, 2013–2014) from the National Commission for Scientific and Technological Research (CONICYT). This work was supported financially by the Biotechnological and Biological Sciences Research Council (BBSRC, United Kingdom) Institute Strategic Programme Grant "Understanding and Exploiting Plant and Microbial Secondary Metabolism" (BB/J004561/1) and the Basal Programme of Conicyt (Chile) for funding of the Centre for Biotechnology and Bioengineering, CeBiB (project FB0001).

Author details

[1]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom. [2]Centre for Biotechnology and Bioengineering (CeBiB), Department of Chemical Engineering and Biotechnology, Universidad de Chile, Beauchef 850, Santiago, Chile.

Received: 25 February 2015 Accepted: 20 May 2015
Published online: 30 June 2015

References

1. Goodfellow M, Kämpfer P, Busse H-J, Trujillo ME, Suzuki K-i, Ludwig W, et al. Bergey's Manual of Systematic Bacteriology, The Actinobacteria, vol. 5. 2nd ed. New York: Springer; 2012.
2. Challis GL, Hopwood DA. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. Proc Natl Acad Sci U S A. 2013;100 Suppl 2:14555–61.
3. Gomez-Escribano JP, Bibb MJ. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol. 2014;41(2):425–31.
4. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 2002;417(6885):141–7.
5. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol. 2003;21(5):526–31.
6. Challis GL. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. J Ind Microbiol Biotechnol. 2014;41(2):219–32.
7. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci U S A. 2011;108(15):6258–63.
8. Shin SC, Ahn DH, Kim SJ, Lee H, Oh T-J, Lee JE, et al. Advantages of Single-Molecule Real-Time sequencing in high-GC content genomes. PLoS One. 2013;8(7):e68824.
9. Okoro CK, Brown R, Jones AL, Andrews BA, Asenjo JA, Goodfellow M, et al. Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. Antonie Van Leeuwenhoek. 2009;95(2):121–33.
10. Busarakam K, Bull AT, Girard G, Labeda DP, van Wezel GP, Goodfellow M. *Streptomyces leeuwenhoekii* sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in *Streptomyces* gene trees. Antonie Van Leeuwenhoek. 2014;105(5):849–61.
11. Rateb ME, Houssen WE, Harrison WTA, Deng H, Okoro CK, Asenjo JA, et al. Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment. J Nat Prod. 2011;74(9):1965–71.
12. Rateb ME, Houssen WE, Arnold M, Abdelrahman MH, Deng H, Harrison WTA, et al. Chaxamycins A-D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. J Nat Prod. 2011;74(6):1491–9.
13. Bonfield JK, Whitwham A. Gap5–editing the billion fragment sequence assembly. Bioinformatics. 2010;26(14):1699–703.
14. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 2011;39(13):e90.
15. Bibb MJ, Findlay PR, Johnson MW. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene. 1984;30(1-3):157–66.

16. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16(10):944–5.
17. Otto TD, Sanders M, Berriman M, Newbold C. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. Bioinformatics. 2010;26(14):1704–7.
18. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99'. 1999. p. 45–56.
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
20. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
24. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
25. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, et al. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 2005;33:W455–9.
26. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
27. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42(Database issue):D206–14.
28. Yang C-C, Huang C-H, Li C-Y, Tsay Y-G, Lee S-C, Chen CW. The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. Mol Microbiol. 2002;43(2):297–305.
29. Wu H, Qu S, Lu C, Zheng H, Zhou X, Bai L, et al. Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. BMC Genomics. 2012;13:337.
30. Weaver D, Karoonuthaisiri N, Tsai H-H, Huang C-H, Ho M-L, Gai S, et al. Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol Microbiol. 2004;51(6):1535–50.
31. Huang C-H, Chen C-Y, Tsai H-H, Chen C, Lin Y-S, Chen CW. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol Microbiol. 2003;47(6):1563–76.
32. Y-r L, Hahn M-Y, Roe J-H, Huang T-W, Tsai H-H, Lin Y-F, et al. *Streptomyces* telomeres contain a promoter. J Bacteriol. 2009;191(3):773–81.
33. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406–15.
34. Zhang R, Yang Y, Fang P, Jiang C, Xu L, Zhu Y, et al. Diversity of telomere palindromic sequences and replication genes among *Streptomyces* linear plasmids. Appl Environ Microbiol. 2006;72(9):5728–33.
35. Pang X, Aigle B, Girardet J-M, Mangenot S, Pernodet J-L, Decaris B, et al. Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome. Antimicrob Agents Chemother. 2004;48(2):575–88.
36. Ian E, Malko DB, Sekurova ON, Bredholt H, Rückert C, Borisova ME, et al. Genomics of sponge-associated Streptomyces spp. closely related to *Streptomyces albus* J1074: insights into marine adaptation and secondary metabolite biosynthesis potential. PLoS One. 2014;9(5):e96719.
37. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–30.
38. Zhong L, Cheng Q, Tian X, Zhao L, Qin Z. Characterization of the replication, transfer, and plasmid/lytic phage cycle of the *Streptomyces* plasmid-phage pZL12. J Bacteriol. 2010;192(14):3747–54.
39. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0–a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 2013;41(W1):W204–12.
40. Mutschler H, Meinhart A. Epsilon/zeta systems: their role in resistance, virulence, and their potential for antibiotic development. J Mol Med (Berl). 2011;89(12):1183–94.
41. Sevillano L, Daz M, Yamaguchi Y, Inouye M, Santamara RI. Identification of the first functional toxin-antitoxin system in *Streptomyces*. PLoS One. 2012;7(3):e32977.
42. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278(5338):631–7.
43. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol. 2012;30(7):693–700.
44. Palaniappan N, Ayers S, Gupta S, Habib E-S, Reynolds KA. Production of hygromycin A analogs in *Streptomyces hygroscopicus* NRRL 2388 through identification and manipulation of the biosynthetic gene cluster. Chem Biol. 2006;13(7):753–64.
45. Lin X, Hopson R, Cane DE. Genome mining in *Streptomyces coelicolor*: molecular cloning and characterization of a new sesquiterpene synthase. J Am Chem Soc. 2006;128(18):6022–3.
46. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, et al. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem. 2008;283(13):8183–9.
47. Barona-Gómez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J Am Chem Soc. 2004;126(50):16282–3.
48. Barona-Gómez F, Lautru S, Francou F-X, Leblond P, Pernodet J-L, Challis GL. Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces ambofaciens* ATCC 23877. Microbiology. 2006;152(Pt 11):3355–66.

Nomenclature

α	in chemistry, is a descriptor that indicates the position of the first atom bonded to a functional group. In a polyketide molecule it is a descriptor that refers to a substituent group (*e. g.* methyl) attached to the α-carbon.
β	in chemistry, is a descriptor that indicates the position of the second atom bonded to a functional group. In a polyketide molecule it is a descriptor that refers to a substituent group (*e. g.* hydroxyl) attached to the β-carbon.
γ	in chemistry, is a descriptor that indicates the position of the third atom bonded to a functional group.
Δ*gene*:: *resistance marker*	in molecular biology, indicates that a gene of interest has been deleted and replaced with another gene that confers resistance to an antibiotic.
μ	in the International System of Units is the prefix micro- that represents one millionth, or a factor of 10^{-6}.
Φ	in molecular biology, refers to a bacteriophage.
°C	in the International System of Units is Celcius degrees, a temperature unit.
C	in chemistry, is a descriptor that indicates carbon position.
C-terminus	in chemistry refers to the carboxy-terminus of an amino-acid, peptide or protein.
cis	in polyketide biosynthesis refers to those protein domains of a polyketide synthase enzyme that are present and integrated into the enzyme.
cm	in the International System of Units, centimetre is a measure unit of length that represents one hundredth of a metre.
D	in chemistry refers to the orientation of a substituent of a chiral carbon *dexter* = on the right.

E — in chemistry, is a stereodescriptor that refers to the *entgegen* = opposite configuration of a double bond in a molecule, in which the groups lie on the opposite side of a reference plane passing through the double bond.

h — in the International System of Units, hour is a measure unit of time that comprises 60 minutes, or 3,600 seconds.

k — in the International System of Units is the prefix kilo- that represents one thousand, or a factor of 10^3.

l — in the International System of Units, liter is a measure unit of volume that is equal to equal to 1 cubic decimetre.

L — in chemistry refers to the orientation of a substituent of a chiral carbon *laevus* = on the left.

m — in the International System of Units is the prefix mili- that represents one thousandth, or a factor of 10^{-3}.

M — in the International System of Units is the prefix mega- that represents one million, or a factor of 10^6. In chemistry, is a unit of concentration of a solution that represents 1 mole of solute per litre of solvent.

$[M - H]^-$ — in chemistry represents negative ion mode.

min — in the International System of Units, minute is a measure unit of time that comprises 60 seconds.

n — in the International System of Units is the prefix nano- that represents one billionth, or a factor of 10^{-9}.

N — in chemistry, is a descriptor that indicates nitrogen position.

N-terminus — in chemistry refers to the amino-terminus of an amino-acid, peptide or protein.

O — in chemistry, is a descriptor that indicates oxygen position.

(*R*) — in chemistry, is a stereodescriptor that refers to the configuration of a stereogenic (chiral) centre. The direction of the substituent atoms attached to a chiral centre, based on the Cahn-Ingold-Prelog rules, is clockwise or *rectus* = right-handed.

R — in molecular biology, this superscript refers to a phenotype that is resistant to an antibiotic.

(*S*) — in chemistry, is a stereodescriptor that refers to the configuration of a stereogenic (chiral) centre. The direction of the substituent atoms attached to a

chiral centre, based on the Cahn-Ingold-Prelog rules, is counter clockwise or *sinister* = left-handed.

S — in molecular biology, this superscript refers to a phenotype that is sensitive to an antibiotic.

trans — in molecular biology it refers to the complementation of a mutant with the wild-type gene, which is provided extrachromosomally. In polyketide biosynthesis refers to a protein domain (commonly acyltransferase) that is absent or not integrated into the polyketide synthase enzyme and acts as a discrete enzyme in polyketide biosynthesis

x — in biochemistry, represents any amino-acid.

Z — in chemistry, is a stereodescriptor that refers to the *zusammen* = together configuration of a double bond in a molecule, in which the groups lie on the same side of a reference plane passing through the double bond.

Abbreviations

2xYT	Two times yeast extract-tryptone
A (domain)	Adenylation
AAC	Aminoglycoside *N*-acetyltransferase
Aac/aac	Acetyl-CoA carboxylase
ABC	ATP-binding cassette
ACP	Acyl carrier protein
AHB	3-amino-5-hydroxybenzoate
AHBA	3-amino-5-hydroxybenzoic acid
AminoDAHP	3,4-dideoxy-4-amino-D-arabino-heptulosonate 7-phosphate
AminoDHQ	5-deoxy-5-amino-3-dehydroquinate
aminoDHS	5-amino-5-deoxy-3-dehydroshikimate
AminoQ	Amino quinate
AMP	Adenosine monophosphate
ANT	Aminoglycoside *O*-nucleotidyltransferases
APH	Aminoglycoside phosphotransferase
AT	Acyltransferase
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
attB	Bacterial attachment site
attP	Phage attachment site
b	Nucleotide bases
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool for nucleotide sequences
BLASTp	Basic Local Alignment Search Tool for protein sequences
bp	Nucleotide base pair(s)

C	Cytosine
CDD	Conserved Domains Database
CIP	Cahn-Ingold-Prelog
CoA	Coenzyme A
CPK	Cryptic polyketide
DEBS	6-deoxyerythronolide synthase
DH	Dehydratase
DHCHC	4,5-dihydroxycyclohex-1-ene carboxylic acid
DMDARSV	27-*O*-demethyl-25-*O*-desacetylrifamycin SV
DMSO	Dimetil sulfoxide
DMRSV	7-*O*-demethylrifamycin SV
DMT	Drug/metabolite transporter
DNA (medium)	Difco nutrient agar
DNA (molecule)	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
ER	Enoylreductase
FAD	Flavin adenine dinucleotide
FDA	Food and Drug Administration
FMN	Flavin mononucleotide
G	Guanine
GalU	UTP:α-D-glucose-1-phosphate uridylyltransferase
GlnA	Glutamine synthase
GltA	Citrate synthase
GUS	β-glucuronidase
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
ISP2	international *Streptomyces* project 2
ISP3	international *Streptomyces* project 3
ISP4	international *Streptomyces* project 4
ISP6	international *Streptomyces* project 6
ISP7	international *Streptomyces* project 7
IT	Ion trap

KR	Ketoreductase
KS	Ketosynthase
LA	Luria agar
LB	Luria broth
LC	Liquid chromatography
m/z	Mass to charge ratio
mA	N^6-methyladenine
mal	Malonyl-CoA
mC	5-methylcytosine
Mce	Methylmalonyl-CoA epimerase
Mcm	(2*S*)-methylmalonyl-CoA mutase
MCS	Multi cloning site
Mct	Methylmalonyl-CoA carboxyltransferase
MFS	Major facilitator superfamily
MIC	Minimal inhibitory concentration
mmal	(2*S*)-Methylmalonyl-CoA
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MT	Methyltransferase
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
neo	Kanamycin resistance gene
NRPS	Nonribosomal peptide-synthetase
nt	Nucleotide
OD	Optical density
PAC	P1-phage derivative artificial chromosome
PacBio RS II SMRT	Pacific Biosciences RS II's Single Molecule, Real-Time
Pcc	Propionyl-CoA carboxylase
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvic acid
Pgm	Phosphoglucomutase
PKS	Polyketide synthase

PP_i	inorganic pyrophosphate
R2	Regeneration 2
R2-S	Regeneration 2 without sucrose
R3	Regeneration 3
R5P	D-ribose 5-phosphate
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
RR	Response regulator
S7P	D-sedoheptose-7-phosphate
SAM	*S*-adenosyl-L-methionine
SDS	Sodium dodecyl sulfate
SET	Sodium, EDTA, tris
SFM	Soya flour-mannitol
SMART	Simple Modular Architecture Research Tool
SMRT	Single Molecule Real-Time
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
STET	Sucrose, Tris, EDTA and Triton X-100
TE	Thioesterase
TES	[*N*-Tris (hydroxymethyl) methyl-2-aminoethanesulfonic acid]
TIC	Total ion chromatogram
ToF	Time of flight
TRIS	Tris(hydroxymethyl) aminomethane
TSB	Tryptone soya broth
TSB-YEME	50 vol% of TSB and 50 vol% of YEME media
UDP	Uridine diphosphate
VRE	Vancomycin-resistant *Enterococcus*
YEME	Yeast extract-malt extract
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronide
XML	eXtensible Markup Language
ZMW	Zero-Mode Waveguide