

Deanship of Graduate Studies
Al-Quds University



جامعة القدس
Al-Quds University

Phylogenetic Typing of *Streptomyces* Species Using
Multilocus PCR Analysis

Mohammad Ahmad Al-Qadi

M.Sc. Thesis

Jerusalem-Palestine

1434Hijri/2013

Phylogenetic Typing of *Streptomyces* Species Using Multilocus PCR Analysis

Prepared By:

Mohammad Ahmad Al-Qadi

B.Sc. Arab American University, Palestine

Supervisor: Dr. Sameer A. Barghouthi

A thesis submitted in partial fulfillment of requirements for the degree of Master of Medical Laboratory Sciences / Microbiology and Immunology Track, Department of Medical Laboratory Sciences, Faculty of Health Professions, Al-Quds University.

1434Hijri/2013

Al-Quds University
Deanship of Graduate Studies
Department of Medical Laboratory Sciences

Thesis Approval

Phylogenetic Typing of *Streptomyces* Species Using Multilocus PCR Analysis

Prepared By: Mohammad Ahmad Al-Qadi
Registration No. 20913084

Supervisor: Dr. Sameer A. Barghouthi

Master Thesis submitted and accepted, Date: 06/01/2013

The names and signatures of the examining committee members are as follows:

- 1- Head of committee: Dr. Sameer A. Barghouthi
- 2- Internal Examiner: Dr. Ibrahim Abbasi
- 3- External Examiner: Dr. Walid Basha

Jerusalem-Palestine

1434/2013

Dedication:

I dedicate my work to those dearest to me, my dear mother who faced a great deal of time and difficulties for the sake of my success and wellbeing, my kind father whose advice have guided my path through difficult times, my lovely sisters who supported my efforts, my late brother with whom I had the best companionship, my supportive uncle Ali who helped me in my duties, my loving wife whose affection and efforts made this possible.

Mohammad Ahmad Al-Qadi

Declaration

I certify that this thesis submitted for the degree of master, is the result of my own research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed-----

Mohammad Ahmad Al-Qadi

Date: 14/12/2012

Acknowledgments

I offer my deepest feelings of gratitude and acknowledgment to my supervisor Dr. Sameer A. Barghouthi for his sincere efforts in my training, supervision, and guidance. The study was conducted in the research lab of the faculty of health professions in Al-Quds University.

Abstract:

Streptomyces species are important bacteria; they produce several bioactive metabolites such as antibiotics including antibacterial, antifungal, antiviral, anticancer, and other secondary metabolites that are important for human health, veterinary medicine, soil and plants. Classification of *Streptomyces* using classical methods; morphological, microscopic, and biochemical, suffers from major weaknesses and confusion. Classical methods are difficult to apply to the large number of available *Streptomyces spp.* (>570), cumbersome, expensive, time consuming, and most importantly are not applicable for routine work, require extensive experience which hinder their application and bacterial identification. With the development of molecular biology, a possible solution to most problems is in sight. 16S ribotyping is being applied, yet it failed to identify species. This study was designed to categorize *Streptomyces spp.* (known and unknown) into smaller and manageable subgroups and simplify identification and classification of the genus *Streptomyces*, the multi-locus gene analysis in addition to the 16S rRNA gene was deemed the method of choice; the selected genes were Superoxide dismutase (SOD), Pyruvate phosphate dikinase (PPD), Sigma Factor (σ f), Amino acid adenylation domain protein (aaa), [2Fe-2S] protein (IS), and Transcriptional regulator (TR). Based on available DNA sequences, BLAST search with published sequences, 17 primers were designed under the following criteria: amplicon sizes must be distinguishable by agarose gels, all primers should have similar melting temperatures (T_m) (60°C), and that no any primer pair will react with all *Streptomyces* species. Worldwide collected soil samples were cultured on citrate-glycerol medium or ISP4, suspected isolates were subjected to PCR in a multiplex setting. SOD amplicon was sequenced for a limited number of isolates in order to validate and confirm the ability of the proposed Multilocus typing to separate species into subgroups. Theoretically, *Streptomyces spp.* can be subdivided into genetic subgroups based on the presence or absence of selected loci. The sequential arrangement of loci had generated a flow chart which placed a given isolate in one of 192 possible subgroups rendering its accurate identification highly possible.

تعد البكتيريا من فصيلة ستريبتومايسز (*Streptomyces spp.*) ذات أهمية كبيرة، إذ تنتج عدة مواد ذات خصائص حيوية هامة كالمضادات الحيوية بما فيها من مواد مضادة للبكتيريا، الفيروسات، الفطريات، ومواد مضادة للخلايا السرطانية، بالإضافة إلى منتجات ثانوية أخرى لها استخدامات و تطبيقات مهمة في مجال الصحة، الطب البيطري، والزراعة.

إن تصنيف الستريبتومايسز (*Streptomyces*) باستخدام الطرق التقليدية، بناء على الطراز الشكلي ، أو المجهرى، وبناء على خصائص الكيمياء الحيوية يعاني من لبس ونقاط ضعف واضحة. فالطرق التقليدية صعبة التطبيق بسبب وجود عدد كبير من أنواع الستريبتومايسز (*S. species*) (< 570)، ناهيك عن تعقيد هذه الطرق وارتفاع كلفتها. كما أنها بحاجة إلى وقت طويل، وهي غير قابلة للتطبيق بشكل روتيني، وهذا لأنها تحتاج إلى خبرة عالية، مما يعيق تطبيقها واستخدامها في التشخيص. والآن، مع تطور علم الأحياء الجزيئي هناك بديل محتمل يلوح في الأفق. وبالرغم من استخدام العامل الريبوسومي (16S rRNA) في التصنيف إلا أنه ما زال قاصرا في التصنيف على مستوى النوع (Species). هذه الدراسة تهدف إلى تقسيم أنواع الستريبتومايسز (المعروفة منها وغير المعروفة) إلى مجموعات صغيرة يسهل التعرف عليها وتصنيفها ضمن جنس الستريبتومايسز (*Streptomyces* Genus). إن استخدام جينات متعددة (multilocus) بالإضافة إلى جين 16S rRNA هي الوسيلة المتبعة في هذه الدراسة، أما بالنسبة للجينات المختارة للدراسة فهي Superoxide dismutase (SOD), Pyruvate phosphate dikinase (PPD), Sigma Factor (σ), Amino acid

adenylation domain protein (aaa), [2Fe-2S] protein (IS), and Transcriptional regulator (TR).

و قد اختيرت بناء على المعلومات المتاحة عن تسلسل القواعد النيتروجينية للحمض النووي DNA في قواعد البيانات المنشورة على الانترنت (Data base)، ومقارنة سلسلة القواعد النيتروجينية المختارة بمتسلسلات القواعد النيتروجينية المنشورة على قواعد البيانات. لقد استخدمت 17 بادئة (primer) في هذه الدراسة صممت وفق المعايير التالية: ان يكون حجم الناتج (amplicon) لكل منها مميزا عن غيره عند عرضه على جل الاجروز (agrose gel)، أن تكون جميع البادئات لها نفس درجة الانفصال (Tm) وهي 60 درجة مئوية، وأيضا أن لا يتفاعل أي زوج من هذه البادئات مع جميع أنواع الستريبتومايسز. تم جمع عينات التربة من مناطق مختلفة من العالم وزرعت على وسط بكتيري محتوي على ملح السترات بالإضافة إلى الجليسيرول (citrate-glycerol) وعلى وسط بكتيري آخر معروف و معد خصيصا لهذه البكتيريا (ISP4). جهزت الأحماض النووية المستخرجة من هذه البكتيريا لتفاعل الأنزيم المبلمر (PCR) باستخدام تقنية الخلط بين البادئات (multiplex PCR). لقد تم استخدام تقنية تحليل القواعد النيتروجينية (sequencing) على عدد معين من العزلات باستخدام جين SOD لتأكيد وإثبات قدرة هذه الطريقة على فصل أنواع بكتيريا ستريبتومايسز إلى مجموعات صغيرة بناء على وجود أو غياب المواقع المختارة (loci) من الجينات.

لقد أنتجت هذه الدراسة مخططا يضع هذه البكتيريا ويقسمها إلى 192 مجموعة (Cluster) مما يزيد من فرص التعرف عليها.

Table of abbreviation

Abbreviation	Full word
aaa	Amino acid adenylation domain protein
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CCG media	Sodium Citrate Casaminoacids Glycerol media
ELISA	Enzyme-Linked Immunosorbent Assay
G7	Golden mixture 7
IE	Indirect ELISA
IS	[2Fe-2S] protein
ISP4	International <i>Streptomyces</i> project 4
LFRFA	Low-frequency restriction fragment analysis
MLEE	Multilocus enzyme electrophoresis
MLSA	Multilocus sequence analysis
SIMCA	Soft independent modeling by class analogy
T _m	Melting temperature
PAGE	Polyacrylamide gel electrophoresis
RAPD	Random amplification of polymorphic DNA
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis

Abbreviation	Full word
QUBC	Al-Quds University Bacterial Collection
QUGP	Al-Quds University General Primers
SDS	Sodium dodecyl sulfate
σ f	Sigma Factor
SOD	Superoxide dismutase
spp.	Species
TR	Transcriptional regulator

Table of Contents

Declaration	i
Acknowledgment	ii
Abstract	iii
Abstract in Arabic الملخص	iv
Table of abbreviation	vi
Table of contents	viii
List of tables	xi
List of figures	xii
Chapter 1: Introduction	1
1. literature review	2
1. 1 the name <i>Streptomyces</i>	2
1. 2 Bases of classification	3
1. 2. 1 Chemotaxonomy	3
1. 2. 2 Serology	4
1. 2. 3 Numerical classification based on mycelium color	5
1. 2. 4 Genotypic methods in taxonomy	6
1. 2. 4. 1 DNA-DNA hybridization.....	6
1. 2. 4. 2 Multi locus typing	7

1. 3 Problem statement.....	8
1. 4 Aims.....	8
1. 5 Hypothesis.....	8
1. 6 Approaches.....	8
Chapter 2: Materials & Methods.....	10
2. 1 Isolation of putative <i>Streptomyces</i> and culture media.....	10
2. 2 Gene selection.....	11
2. 2. 1 Multiplex mixtures.....	13
2. 3 Calculation and taxonomic clustering.....	13
2. 3. 1 Cluster numbering.....	15
2. 4 Primer design.....	16
2. 4. 1 Specificity of primers.....	17
2. 5 DNA extraction, amplification & sequencing.....	17
2. 5. 1 DNA extraction.....	17
2. 5. 2 Amplification.....	18
2. 5. 3 Sequencing.....	18
Chapter 3: Results.....	19
3. 1 Diversity of <i>Streptomyces</i> species.....	19
3. 2 Specificity of the primers.....	22

3. 3 Screening of selected isolates with all primer pairs applied as multiplexes for clustering	22
3. 4 16 S rDNA groups (A, B and C)	22
3. 5 Other loci	25
3. 6 Assignment of isolates to specific cluster	26
3. 7 Construction of a flow chart	27
3. 8 Superoxide dismutase amplicon sequencing and generation of phylogenetic tree	29
3. 9 One soil sample	30
3. 10 Distribution of tested isolates among hypothesized multiplex clusters	31
Chapter 4: Discussion and Conclusions	32
Recommendation	34
References	35
Appendix	40

List of Tables

Table 2.1: Forward and reverse primers of the different loci and the 16S specific and general primers used in this study for amplification and sequencing.....	12
Table 2.2: Multiplexes prepared for testing the different putative <i>Streptomyces</i> isolate.....	13
Table 2.3: grouping of <i>Streptomyces</i> isolates based on 16S rDNA gene.....	15
Table 3.1: Summary of all tested isolates showing the presence or absence of specified locus.....	26
Table 3.2: Five putative isolates were obtained from the same soil sample.....	30
Table 3.3: Assigned isolates to the hypothesized generated clusters	31
Table A1: Size of the PCR product in the two sets of 16S RRNA.....	41

List of Figures

- Figure 2.1: Two different *Streptomyces* species; QUBC101 photographed under 4X light microscopy showing the typical thin mycelia; and QUBC102 photographed under 1x dissecting microscopy showing the fungus-like appearance.....10
- Figure 2.2: Molecular loci flowchart of *Streptomyces* PCR reactions based on the presence or absence of loci as detected by PCR amplicons.....14
- Figure 3.1: BLAST results showing poor homology between claimed *Streptomyces species* that varies from <1% to >99% indicating poor taxonomy of the current Genus *Streptomyces*.....20
- Figure 3.2: Super Oxide Dismutase (SOD) BLAST results showing major variations among different *Streptomyces* spp. yet consistent within each species. Some inconsistencies suggest that such species may have been misidentified; red arrows.....21
- Figure 3.3A: Three isolates belong to 16s-A group, QUBC 259 belongs to 16S-B since only one band is showing in the first reaction of 16s reaction.....23
- Figure 3.3 B: Three isolates Showing QUBC268 (16S-B); QUBC 315 (16S-A), and QUBC 261 (16s-C) yet all three isolates share the same profile (2mix) with no bands.....24
- Figure 3.3 C: Three additional isolate showing QUBC 269 (16s-C) and QUBC 313 and 314 (16S-B).....24
- Figure 3.4: Flowchart of the 16S-A group.....27
- Figure 3.5: Flowchart of the 16S-B group.....28
- Figure 3.6 : Flowchart of the 16S-C group.....28
- Figure 3.7: Superoxide dismutase amplicon (333bp) was sequenced (QUBC isolates, with their cluster number indicated, additional published (SOD) sequences (scientific binomial names) were incorporated to construct this phylogenic tree.....29

Figure 3.8: Two or three different morphological isolates from one soil sample (Germany) on the CCG medium.....	30
Figure A1: <i>Streptomyces coelicolor</i> 16S rRNA and locations of all 16S primers.....	40
Figures A2: Loci used in this study showed discriminatory power among published genomic sequences of <i>S.</i> species.....	41
Figure A2.1: Superoxide dismutase from <i>Streptomyces coelicolor</i> A3(2), it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	42
Figure A2.2: Sigma factor from <i>Streptomyces scabiei</i> 87.22, it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	43
Figure A2.3: [2Fe-2S]-protein GBAD from <i>Streptomyces scabiei</i> 87.22, it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	45
Figure A2.4: Pyruvate phosphate dikinase from <i>Streptomyces coelicolor</i> A3(2), it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	47
Figure A2.5: Amino Acid Adenylation Domain Protein from <i>Streptomyces flavogriseus</i> ATCC 33331, it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	50
Figure A2.6: Transcriptional regulatory from <i>Streptomyces coelicolor</i> A3(2), it will allow sequencing since it is widely distributed among <i>Streptomyces</i> spp.....	52
Figure A2.7: rDNA (16 S) from <i>Streptomyces coelicolor</i> A3(2), it will allow sequencing since it is house keeping among <i>Streptomyces</i> spp.....	54

Chapter 1

Introduction

Streptomyces are filamentous soil bacteria that are gram positive with high G+C content; the bacteria are similar to fungi in having hyphae, mycelia, and spores. Furthermore, *Streptomyces* are mostly saprophytic organisms (1, 2, 3). The importance of *Streptomyces* is due to their ability to produce important bioactive metabolites such as antibiotics and antiparasitic agents. For example, *S. avermitilis* produces the antiparasitic compound ivermectin (1, 2, 3). Other important concerns include the pathogenicity to plants; *Streptomyces scabies* is widely pathogenic to potato (4). Others cause human infections (5, 6, 7, 8). In one of the studies, out of 92 isolates only two isolates (*S. griseus* and *S. somaliensis*) were subjected to complete species identification (8). *Streptomyces* are an uncommon cause of infection in human, but they may cause superficial disease after traumatic inoculation. They also cause mycetomas in the feet of farmers and perianal soft tissue infection (6, 7). Other less familiar visceral diseases are seen in HIV (Human immunodeficiency virus) and immune compromised people (5, 8). Classical or traditional methods for the identification and classification of *Streptomyces* are available based on morphological, biochemical, nutritional, structural, and physiological characteristics of the bacterium (9, 10, 11). These methods have limited success and applicability due to noncultivable organisms or due to some variable morphology; a species may fit into different patterns confusing the applicability of these methods (9, 12). In addition, descriptive and morphological traits and colony features are subjective, and may be interpreted differently by different researchers. Phenotypic and biochemical characterization suffer from being incapable of resolving species, from low reproducibility, time consumption, and being labor intensive, as well as requiring expertise in *Streptomyces* taxonomy making them less practical (11, 13, 14). Accordingly, another system or way to diagnose and identify this microbe is required. Molecular methods and bioinformatics represent a strong candidate system to settle this issue.

Starting with the 16S r RNA gene is essential since it is well established that this gene can define the genus and in some cases the species. On the downside, being highly conserved, 16S cannot discriminate between closely related species. Again the taxonomy of the bacteria

in the concept of Phylogenetic (16S ribotyping) is acceptable from domain to genus, but at the species level is not always possible. Microbial taxonomy is not dependent on the evolutionary process to speciation, hence the use of 16S rRNA gene sequence alone is not sufficient, but may be used with other genomic information (phylogenomic concept), multiple loci, and sequence analyses (MLSA) in applying phylogenomics which is based on more than one gene up to chromosomal and genome level analysis (15).

1. Literature review:

1.1 The name *Streptomyces*:

The 1940s was the time in which it has been well known that anaerobic pathogenic species of *Actinomycetes* cause Actinomycosis in human being and lumpy jaw in animals (16, 17). First time description of *Streptomyces* established that the organism is aerobic, forming spores - bearing aerial mycelia. These findings initiated the new organism; Actinomyces are anaerobic species (16). The number of reported species is large, so they need to be reclassified since they encompass a heterogeneous group of organisms when subjected to BLAST analysis (9, 14).

Waksman and Henrici (1943) classified aerobic, aerial spore bearing mycelia microorganisms based on the morphological descriptions and identification keys. Logically, Waksman and Henrici found that *Nocardia* is a valid name that describes aerobic, fragmenting, non-sporulating types of *Actinomycetes* (16, 17). Waksman and Henrici provided the name for the family to be *Streptomycetaceae*. The cause of this family name to become logically accepted with botanical code recommendations, which recommend that family name be derived from generic name. Waksman and Henrici give example of species that are colorless in organic or synthetic media, proteolytic (liquefying gelatin), peptonizing the milk with an alkaline reaction and produce an earthy smell. Accordingly, the name *Actinomyces albus* was replaced with the new name *Streptomyces albus* (16).

1.2 Bases of Classification

1.2.1 Chemotaxonomy:

Fatty acid analysis has been used as a tool to assign an isolate to a particular genus (18). It has also been used for species delimitation of *Bacterioides* (19), and in strain identification of *Corynebacterium* (20).

Saddler (1987) analyzed degraded fatty acid profiles of *Streptomyces cyaneus* by gas chromatography. The data were analyzed by Soft independent modelling by class analogy (SIMCA) statistical package and resulted in three clusters; the largest contains 16 strains, the second cluster contains the rest of *S. cyaneus* strains along with other *Streptomyces* spp.. The third cluster contains a number of blue-spored soil *Streptomyces* (21).

Whole cell analysis is another chemotaxonomic tool that can be useful in *Streptomyces* taxonomy. Whole cells are thermally degraded and examined using mass spectroscopy, resulting in a fingerprint that can be used to classify an isolate (22).

The use of rapid biochemical testing in the taxonomy of *Streptomyces* was attempted by Goodfellow. The study utilized a fluorogenic probe -substrate conjugate. The target enzyme is revealed the presence of the enzyme. The strains used in the study were tested for 14 enzyme activities. The study produced reproducible data, but showed incoherent results in all test strains, which limited the usefulness of biochemical tests as an independent taxonomic tool (23).

Actinophages were used also for genus and species identification. Polyvalent *Streptomyces* phages were studied by Chater in 1986, species specific phages are. Most actinophages are useful in genus delimitation, however, species specific are less useful than polyvalent actinophages (9, 24).

Generating protein profiles of whole cells of *Streptomyces* can be used in taxonomy of unknown *Streptomyces* at the species and subspecies level. The resultant complex protein banding pattern is visualized by electrophoresis on Polyacrylamide gel. A study focused on

protein profiles of 32 strains of *Streptomyces* and 5 *Streptoverticillium* strains by SDS-Polyacrylamide gel electrophoresis (PAGE) demonstrated the capacity of protein profiles to serve taxonomy (25). Indeed, PAGE was applied to classify Streptomycete, which is known to cause potato scab (26). Multilocus enzyme electrophoresis (MLEE) is another method for protein profiling. This method was applied in a study by using 11 different enzymes and viewing protein profiles from polyacrylamide gels to differentiate *Streptomyces spp.* to each other (27).

Paper chromatography can be used in the identification of the *Streptomyces spp.* Paper chromatography is applied to whole cell lysate of the bacterial isolate; this method can discriminate between *Nocardia* and *Streptomyces* based on variation in mesodiaminopimelic acid. The justification for this approach is that the two genera; *Nocardia* and *Streptomyces* share morphological similarities (28).

1.2.2 Serology

Ridell and his colleagues (1986) used Ouchterlony double diffusion method to study *Streptomyces* and to distinguish between *S. lavendulae* and *Streptoverticillium*. The study also revealed some relation between *N. dassonvillei* and *S. griseus* on one hand and *S. flavopersicum* on the other (29).

Kirby and Rybicky (1986) used antisera along with indirect ELISA (IE) as a taxonomic tool to identify and classify *S. amulatus* and *S. cattleya*; in contrast to other methods (Ouchterlony double diffusion , nucleic acid hybridization, morphology, and pigment production) based identification. The study claims that indirect ELISA is more sensitive, more quantitative, easier to read, and less time consuming (30). However, serological methods require prior preparations of antisera to individual species that renders them obsolete for requiring physical storage of an array of antisera. Replenishing antisera supplies and expired lots will undoubtedly introduce unnecessary variations, some of which may not be predictable.

In a try to improve on the previous serological methods mentioned, Wipat and his colleagues (1994) produced monoclonal antibodies against the spores of *S. lividans*. The antibodies produced were tested for *S. lividans* specificity, and proved to be specific to cluster 21 of Williams grouping including *S.lividans* (31).

1.2.3 Numerical classification based on mycelium color

One of the most important attempts in the numerical taxonomy of *Streptomyces* was conducted by Williams. A total of 475 strains were analyzed for their phenotypic characteristics like spore chain morphology and pigmentation, antimicrobial activity, biochemical tests, and antibiotic resistance patterns; 139 characters were investigated. The resulting clusters included single strain clusters, minor strain clusters consisting of less than five strains, and major clusters containing up to seventy one strains. Single strain clusters, minor clusters, and homogeneous clusters were considered as consisting of one species. Members of major clusters were regarded as species groups due to their highly diverse and heterogeneous characteristics, until further chemical and genetic analysis proves the same or otherwise. The similarities between clusters were assessed using a standard resemblance coefficient. Williams pointed to the value of numerical taxonomy as an important tool along with chemical, genetic and serologic analysis. The study also recommended that the genus *Streptomyces* be expanded to include species that bear single or short chains of spores on their substrate mycelia. Members of the family *Streptomycetaceae* were also excluded from the *Streptomyces* genus, and showed very little similarity to the genus members (32).

Williams *et al.* (1983) was revised by Kampfer using 329 tests and 821 strains including ones that were not included in the Williams study (9). A later study conducted also by Williams *et al.* (33) produced a database for the identification of *Streptomyces* using the Willcox Coefficient (34).

In a study carried out by Ismail *et al.* (2006) at the Suez University, Egypt, fifteen *Streptomyces* isolates were analyzed depending on a morphologic and biochemical basis to generate numerical profiles for each of the studied isolate (35). The study utilized 53 character units regarding mycelium color, growth on malt yeast extract agar, growth on inorganic salts starch agar, pigment formation, morphology of spore chain, spore surface, reduction of nitrite, hydrolysis of starch, production of H₂S, decomposition of cellulose, liquefaction of gelatin, peptonization and coagulation of milk, utilization of (citrate, different nitrate and nitrite salts, urea, peptone, and olive oil). Each character was scored as a code (0 or 1) based on its presence or absence. An isolate profile is then represented by a series of zeros and ones. The resulting profile codes were analyzed using the Canberra distance Coefficient to estimate the

similarity percentage between paired isolates. The numerical analysis results revealed a high similarity percentage between isolates of the same mycelium color. The study concluded that mycelium color might be a good basis for the classification and taxonomy of *Streptomyces*. It also concluded that numerical taxonomy is a useful tool in *Streptomyces* classification (35). However, numerical taxonomy suffers from instability of morphological behavior of *Streptomyces* spp. and the large number of tests that must be performed for each isolate. The method dictates that an established list or keys encompassing all *Streptomyces* spp. must be available for identification purposes.

1.2.4 Genotypic methods in taxonomy:

Genotypic analysis has contributed greatly to our knowledge about *Streptomyces* taxonomy, as genes are relatively not affected by growth conditions, medium composition, or environmental factors (9).

Streptomyces differ from most other eubacteria in two aspects; they have large (9×10^6 bp) and linear genomes. (*Streptomyces avermitilis* MA-4680 DNA, complete genome GenBank: BA000030.3)

1.2.4.1 DNA-DNA hybridization

DNA-DNA hybridization is one tool used to identify the species of an unknown isolate. The hybridization of single stranded bacterial DNA with an artificially made DNA sequence is examined to determine percent homology. In a study done in 1986 by Mordarski, the groupings assembled by phenetic and taxonomic studies were evaluated. The strains that make the *Streptomyces albidoflavus* cluster were divided into three groups, as done by numerical and phenetic studies, but although one subcluster matched its counterpart in the previous studies, the other two did not match with the rest of the groupings. Another study that aimed to compare genetic grouping and phenetic grouping was carried in 1996, and concluded also a discrepancy between genetic and phenetic grouping (9). In 1990 one study utilized DNA-DNA hybridization to unify the genus *Streptoverticillium* with the genus *Streptomyces* (36).

Low-frequency restriction fragment analysis (LFRFA) is a molecular method of analysis in which a genome is cut at different points by a restriction endonuclease and visualized by pulsed-field gel electrophoresis (PFGE). LFRFA was used to classify 59 strains belonging to 8

Streptomyces species, and demonstrated some usefulness in this field. However, LFRFA studies have shown some discrepancies in which interspecies relationships were poorly resolved and contradictory. Therefore, its usefulness is limited to finding very closely related strains. (37).

Another molecular method; randomly amplified polymorphic DNA (RAPD) PCR uses a single primer of random sequence to produce a PCR product with a distinctive fingerprint. This method requires stringent standardization to obtain reproducible results, otherwise inconsistent results and absence of recognizable PCR products were observed as in a study done by Mehling *et al.*, in 1995 examined the relationship between *S. virginiae* and *S. lavendulae* using RAPD, and compared it with previous phenotypic and genotypic studies (38, 39).

Specific gene sequencing is a particularly useful tool in *Streptomyces* taxonomy, as some genes are highly conserved (40, 41). Stackebrandt found three regions; α , β , and γ in the 16S rRNA gene that have sufficient variation to be used in genus and species identification of *Streptomyces* (40, 41). Other genes that have been found useful in *Streptomyces* taxonomy include 23S rRNA, 5S rRNA genes, tryptophan synthase gene, and housekeeping genes (39, 42, 43).

1.2.4.2 Multilocus typing

Multilocus typing is a phylogenetic concept achieved by studying five to eight genes to explore relatedness among bacterial species and strains (15). Multilocus typing has been explored by several researchers working with different bacteria; Liu *et al.*, 2009 (44) working with *Escherichia coli* O157:H7 described subtyping highly conserved genes to be useless due to the lack of sequence variation found within analyzed housekeeping and virulence genes, which prevented subtyping of *Escherichia coli* O157:H7. (*rhs*) Genes Are Potential Markers for Multilocus Sequence Typing of *Escherichia coli* O157:H7 Strains (44). Guo 2008, have utilized four housekeeping genes and DNA sequencing to classify phytopathogenic *Streptomyces* spp. (14).

In the present multilocus study, five genes were tested that did not belong to the housekeeping genes in addition to sigma factor gene and three differential primers of the 16S rRNA gene.

In this study, a new approach named Multilocus Numerical Classification was explored.

1.3 Problem statement:

There is no simple, inexpensive, and robust method to identify and classify *Streptomyces* spp, except for genome sequencing which may not be practical for routine work and may not be accessible to all.

1.4 Aims:

The aim of this study is to find a simple, inexpensive, quick, and practical method for the classification of *Streptomyces* into taxonomic clusters.

1.5 Hypothesis:

Current identification of *Streptomyces* species appears to be suffering major problems and confusion. Nucleotide data bases can be searched to find genes, which together with the 16S rRNA gene can discriminate and classify the genus *Streptomyces* into smaller and manageable taxonomic clusters.

1.6 Approaches:

Dividing the current members of the genus *Streptomyces* into a sufficiently large number of clusters (~200) using a selected number of discriminatory genes/loci (7 or more) that can discriminate *Streptomyces* from other genera and *S. species* from each other by producing different PCR-molecular patterns suitable for *Streptomyces* typing. When amplified, these loci should produce a number of specific patterns (clusters), each pattern will represent a cluster that is sufficiently small, encompassing a limited number of species (three species per cluster for a total of ~600 species). Species within each cluster can then be resolved based on additional features or tests. The rRNA gene and six other selected loci should produce a taxonomic key (a flow chart). The designed primers must be specific for *Streptomyces*, but must be present in a limited fraction of the all *S. species* (this is a hypothetical assumption that must be met). Availability of genome sequences in the different nucleotide data bases should allow testing and clustering based on the hypothesized multilocus approach. Clustering

Streptomyces spp. will simplify typing and identification; therefore resolving the predicament of identifying each *Streptomyces* spp. from among >576 known species. In addition, the system can be expanded to accommodate newly discovered *S.* species; if 200 or 400 new species are to be discovered that were randomly distributed among clusters, then each cluster should contain 4 or 5 species which still can be easily identified within their cluster.

Chapter 2

Materials & Methods

2.1 Isolation of Putative *Streptomyces* and Culture Media:

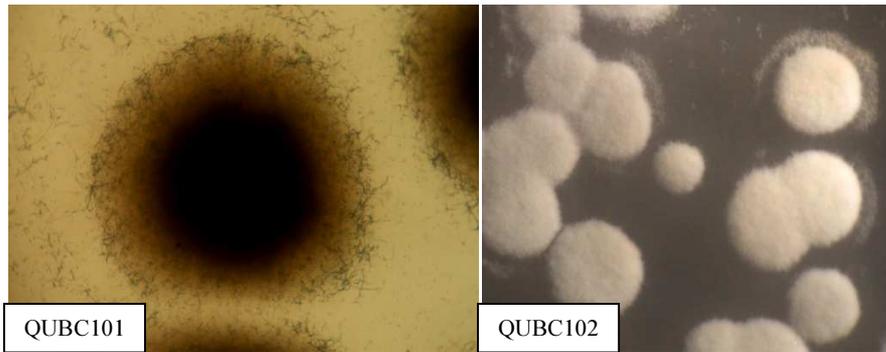


Figure 2.1: Two different *Streptomyces* species; QUBC101 photographed under 4X light microscopy showing the typical thin mycelia; and QUBC102 photographed under 1x dissecting microscopy showing the fungus-like appearance. QUBC102 was identified as *Streptomyces griseusref* since it shared 98.9% of a partial 16s rDNA amplicon, it was isolated from a pigeon feather identified by Barghouthi (49). Both had distinct odder typical of *Streptomyces*.

Figure 2.1 shows two cultured isolates that were part of this study at Al-Quds University. QUBC101 was grown on (Sodium Citrate Casaminoacids Glycerol media (CCG)) and photographed with a digital camera under 4x magnification of bright field microscope. QUBC102 on (CCG) was photographed under dissecting microscope (1x; final magnification was 10 times higher).

All bacterial isolates were cultivated from around twenty worldwide collected soil samples to obtain representative samples and to reduce repetitive isolation of the same isolate or species. The soil samples (200-300 g) were collected from a depth of five centimeters in fresh plastic bags and stored at room temperature.

The medium used for isolation is a citrate-glycerol (CCG) medium which was formulated by (Barghouthi and Ayesh) and that contained caseaminoacids, sodium citrate, Sodium Chloride, cyclohexamide, and glycerol (45). The commercially available medium (ISP4) was mainly used for routine maintenance and purification. ISP4 was modified; it was utilized at one-half strength while maintaining agar concentration at 20%. The cooled autoclaved medium was supplemented with 100 μ g of cyclohexamide/ml. Soil samples (15 mg) were randomly taken

on a pre-flamed spatula and sprinkled evenly over the agar surface, cultivated at 25 °C for up to 3 weeks until optimal growth was obtained. CCG or half-strength ISP4 plates were used to subculture or maintain cultures. Plates were stored refrigerated before and after culturing.

2.2 Gene Selection:

Chromosome mapping for SOD, as in figures 3.1 and 3.2, was used from the available data in gene banks such as NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), in addition to the 16S rRNA gene, other genes selected for this study based on being less conserved than the 16S gene and appeared to be widely distributed, but not present in all *Streptomyces* species. These produced positive and negative BLAST search results with available *Streptomyces* DNA sequences. (Reactive or positive +; and nonreactive or negative -).

Table 2.1: Multilocus Primers. Forward and reverse primers of the different loci and the 16S specific and general primers used in this study for amplification and sequencing.

Locus	F/R	Primer Sequence (5'→3')	-mer	Amplicon size (bp)	<i>T_m</i> (°C)
Superoxide Dismutase	F	5'-GCAGGTAGAAGGCGTGCTCCCA-3'	22	333	63
	R	5'-CCTCGCGTTCCACCTCTCCGG-3'	21		
[2Fe-2S] Protein GBAD	F	5'-ATGTGGTGCTCGCTGGGCACCAG-3'	23	474	61
	R	5'-CATGGAGTGCTATCACTGCGCG-3'	22		
Transcriptional Regulatory	F	5'-AGCAGGACGAAGGTGTCCCGGTC-3'	23	607	62
	R	5'-GCCATCAACGACCGGCTCGCCCT-3'	23		61
Pyruvate Phosphate Dikinase	F	5'-GGTCCATCTGCTCGCGCGG-3'	19	655	63
	R	5'-GGAGTCGAGACCCGTGTCCGAAA-3'	23		
Sigma Factor	F	5'-ACTCGACGATCTGCACCCAGTCG-3'	23	709	63
	R	5'-AAGGGCTGGCTGGTCACCGT-3'	20		
Amino Acid Adenylation Domain Protein (aaa)	F	5'-TGCCGCCAGTTCGAAGAAGTCGTC-3'	25	752	65
	R	5'-GGTGAACATGTTCCGCATCACGGAGAC-3'	27		64
16S RRNA FW1	F	5'-GGTCGAAAGCTCCGGCGGTGAA	22	934	60
	R	5'-GAGTCCCCAACACCCCGAAGG-3'	21		60
16S RRNA FW2	F	5'-CCCTTCACTCTGGACAAGCCC	22	1005	61
	R	5'-GAGTCCCCAACACCCCGAAGG-3	21		60
QUGPRn2*	R	5'-TGACGGGCGGTGTGTACAAG-3'	20	See Table A1	65
QUGPRn3*	R	5-GGCGTGGACTACCAGGGTATC-3	21		63

2.2.1 Multiplex mixtures

Table 2.2: Multiplexes prepared for testing the different putative *Streptomyces* isolates

Multiplex	Composition of Multiplex
First set (4 mix; 4 loci)	Superoxide dismutase, Pyruvate phosphate dikinase, Sigma Factor, and Transcriptional regulator
Second set (2Mix; 2 loci)	Amino acid adenylation domain protein and [2Fe-2S]GBAD protein
First 16S rDNA Set	Fw1 Fw.R 16S rDNA
Second 16S rDNA Set	Fw1 Fw.R Rn2 Rn3 16S rDNA

2.3 Calculations and taxonomic clustering:

Calculations showed that having two 16S, *Streptomyces* specific, forward primers used with two 16S general reverse primers, will produce three 16S rDNA subgroups since the presence of the two forward primers generated similar size PCR products that cannot be distinguished in most cases especially when one is absent. The six selected gene loci, coding the potential patterns and the construction of a dendrogram should separate species; this system with 6 genes and three 16S rDNA groups, has the capacity to accommodate 192 clusters (3x64) (species are ~580 spp.). This process will alleviate the need for sequencing and BLAST analyses. Moreover, the analysis of the patterns produced will be based on a flowchart similar to that illustrated in figure 2.2 below.

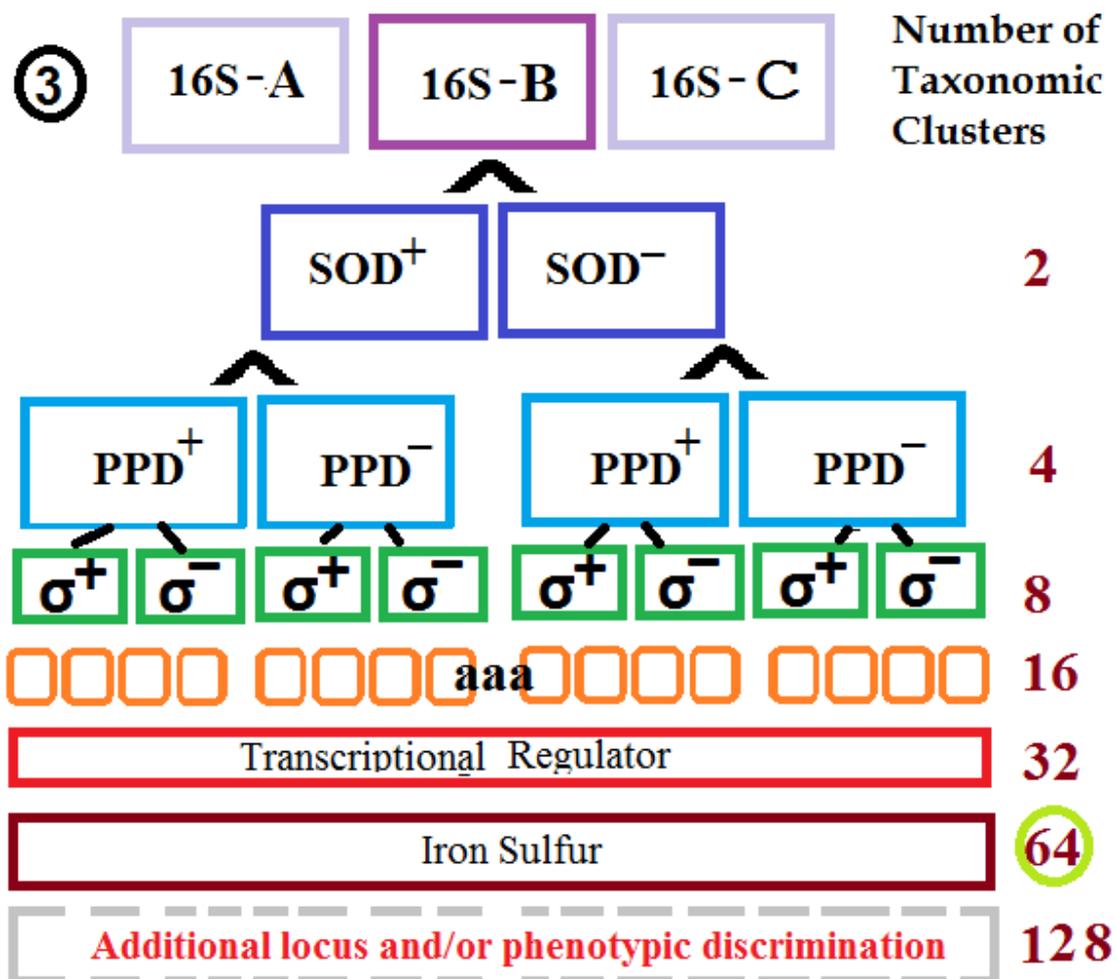


Figure 2.2: Molecular loci flowchart of *Streptomyces*. PCR reactions based on the presence or absence of loci as detected by PCR amplicons. SOD: Superoxide dismutase, PPD: Pyruvate phosphate dikinase, σ : Sigma Factor, aaa: Amino acid adenylation domain protein.

The flowchart shows the order of selected loci used to assign *Streptomyces* to its relevant cluster. In this chart, the first row represents the three possible profiles for rDNA amplicon(s); 16S-A, 16S-B, and 16S-C as detailed in table 2.3. Sod represent the super Oxide dismutase locus. PPD; is the pyruvate phosphate kinase locus followed by sigma factor (σ), Amino acid adenylation domain protein (aaa) and the iron sulfur gene Glycine Betaine Aldehyde Dehydrogenase (GBAD). 64 clusters for each rDNA pattern can be generated bringing the total to 192 potential clusters for all three rDNA groups (A, B, and C).

2.3.1 Cluster numbering:

Table 2.3: grouping of *Streptomyces* isolates based on 16S rDNA gene.

16s multiplex	A Group	B Group	C Group	C Group (excluded)
First set: Fw1 Fw.R 16S rDNA	One band	Two bands	negative	negative
Second set: Fw1 Fw.R Rn2 Rn3 16S rDNA	One band or two	One band or two	Positive	negative

If the two sets of ribosomal primers (Fw1 Fw2. R 16S rDNA and Fw1 Fw2.R Rn2 Rn3 16S rDNA) were positive (amplify) with one band for each set of primers (**A group**)

If the two sets of ribosomal primers (Fw1 Fw.R 16S rDNA and Fw1 Fw.R Rn2 Rn3 16S rDNA) were positive with two bands for the first set of primers and a positive (one or two bands) band at second set of primer (**B Group**)

Positive for the second set of primers and negative for the first set of primers (**C Group**)

If negative (no amplify) for both primers then (**D Group**), such isolates will be excluded until further confirmation is obtained, e.g. superoxide dismutase or another locus sequencing.

The key below illustrates the coding process, once a code is obtained, the code is then assigned to one of the serially arranged clusters (1 to 192: 1-64 belong to A group, 65 -128 belong to B group, and 129-192 belong to C group).

If one or more genes failed to discriminate species, additional or alternative loci can be utilized.

Sequence of any PCR amplicon may help in matching dendrogram identified species with available DNA sequence data of known species and to solve some of the problems that may face this new (dendogramming) approach of typing.

Superoxide dismutase, Pyruvate phosphate dikinase, Sigma factor will be specific to *Streptomyces*, yet, capable of distinguishing the different species (>576 spp. in addition to unknown spp) can be obtained either from published literature or by other methods, [2Fe-2S] protein, Transcriptional regulatory unit, and glycine betaine aldehyde dehydrogenase (GBAD) were selected from the available sequences data bases in addition to the three ABC-16S rDNA variants

(<http://www.ncbi.nlm.nih.gov/genomes/genlist.cgi?taxid=2&type=0&name=Complete%20Bacteria>) (46). It is important to indicate that the selected genes were chromosomal loci avoiding any plasmid encoded genes. Then, data bases were searched using BLAST to determine the matching of a selected gene/locus to a fraction but not all available *Streptomyces* species (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) (47).

Superoxide dismutase sequence as shown by BLAST at the data base found in 27 different *Streptomyces* and some other related bacteria, Pyruvate phosphate dikinase found in 36 different *Streptomyces*, Sigma Factor found in 20 different *Streptomyces*, [2Fe-2S'GBAD] gene found in 7 different *Streptomyces*, and the Transcriptional regulator was found in 17 different *Streptomyces*. All BLAST data indicate that these genes are distributed independently among *Streptomyces* spp. at least at the level of nucleotide sequence and not at the functional level that may or may not be provided by another gene.

Gene Sequences from GenBank: (accession Numbers or gene ID, for each locus shown in BLAST results; See appendix)

2.4 Primer Design:

Table 2.1 shows the details of loci, their primers; sequence, length, and T_m as well as the predicted size of each predicted amplicon. (http://www.promega.com/techserv/tools/biomath/calc11.htm#melt_results) and base-stacking calculated T_m (48).

Each PCR amplicon should specifically identify its corresponding locus and allowing multiplex mixture(s) to be utilized in testing each isolate.

2.4.1 Specificity of primers:

The specificity of each primer pair was tested to make sure that selected primers did not react with other bacteria that may contaminate the process. The exact work procedure in this experiment was done by randomly collecting Gram positive and Gram Negative bacteria collection which were handled exactly as *Streptomyces* isolates and DNA extracted in the same extraction procedure and analyzed to show the lack of cross reactivity.

2.5 DNA extraction, amplification and sequencing:

2.5.1 DNA extraction:

was done according to Barghouthi 2011 and Barghouth and Al Zughayyar 2012 (49,50); colony from agar plate was collected by scraping growth with a sterile scalpel and washed in 350 μ l (J-buffer containing 0.5 μ l / ml of 2-mercaptoethanol), after that the samples were micro centrifuged for one minute at 13000 RPM, then treated with 100 μ l lysozyme and incubated for 30 minute at 37 C°. After removal of lysozyme by centrifugation, 25 μ l of 1% SDS (sodium dodecyl sulfate) and 25 μ l of 0.5 N NaOH were added to lyse the cells and release DNA (51). Finally, the clear supernatant was diluted 15-folds and boiled for 10 min. for more details about DNA see Barghouthi and Al Zughayyar 2012 (Barghouthi SA and Dima K. Al Zughayyar 2012. Detection of *Neisseria meningitidis* and unknown *Gammaproteobacteria* in cerebrospinal fluid using the two-step universal method. [\[Abstract\]](#) [\[Full Text\]](#) [\[Full Article - PDF\]](#) African Journal Microbiol Research pp. 3415-3424. DOI: 10.5897/AJMR11.1213

2. 5. 2 Amplification:

Standard PCR method was optimized. The reaction conditions were as following: initial denaturation by hot start at 95 °C for 2 min, followed by 31 cycles of denaturation at 94° C for 90 s, annealing for 30 s at each 62° C, 60° C, 58° C (49, 52) respectively and extension at 72°C for 105 s. A final extension at 72°C for 2 min. using commercial master mix from Promega Biochemical Company, which contains Taq DNA polymerase, dNTPs, MgCl₂ and reaction buffers at (2X) concentrations. Reaction products were electrophoresed in a 1.5% agarose gel stained with ethidium bromide and were visualized and photographed under UV light (49).

2. 5. 3 Sequencing:

Selected amplicons of Superoxide dismutase gene were extracted using available commercial Kit and sequenced (Bethlehem University), followed by BLAST alignment for species identification. Sequencing of Superoxide amplicon was done to compare sequencing based grouping to Multilocus Generated Clustering. They should be different, but multilocus clusters can be related to the actual sequence of a selected locus, superoxide dismutase or another locus in case of absent superoxide amplification. The Phylogenetic tree was drawn for the sequenced amplicon and multiple alignments using Bioedit software (<http://www.mbio.ncsu.edu/BioEdit/page2.html>) (53). Then the cluster number and isolate number were indicated for every sequenced isolate on the phylogenetic tree. The relationships of the isolates to each other and members of the same cluster can be observed.

Chapter 3

Results

3.1 Diversity of *Streptomyces* species

As indicated by the BLAST generated comparison (Figure 3.1 and 3.2). Figure 3.1 and 3.2 show the BLAST search of *S. griseus* (ADFC02000002.1) against some of the available sequences indicating a similarity that varied from near zero to 96% or higher. These BLAST results indicate that current classification and identities of *Streptomyces* spp. do not reflect their genetic relatedness. The poor identity (<1%) among species of the same genus is vividly clear when chromosomal alignment of a selected species was attempted against other *Streptomyces* spp.

Similarly, BLAST results with superoxide dismutase from *Streptomyces griseus* strain DSM 40931 (Figure 3.2) showed 87% similarity with SOD of another claimed *S. griseus* strain DSM 40660 and 99% similarity to *S. limosus* strain CB143, 91% to *S. cinnamoneus* strain DSM 40005, and 90% to *S. violaceoruber* strain DSM 40049 indicating the poor alignment between members of the same genus or species (*S. griseus*; 87%), whereas better alignment to a different species; *S. limosus* (99%). The poor similarity is indicative of poor taxonomy used to define these strains. Figure 3.2 illustrates that several species were only 9% or less similar to other *Streptomyces* spp.

The results indicate the need to reject the current classification since a genus should only encompass similar members that can be divided into species. The current genus of *Streptomyces* represents a genetically diverse group of organisms that may be divided into several Genera. The current attempt of multilocus clustering and analysis is amply justified.

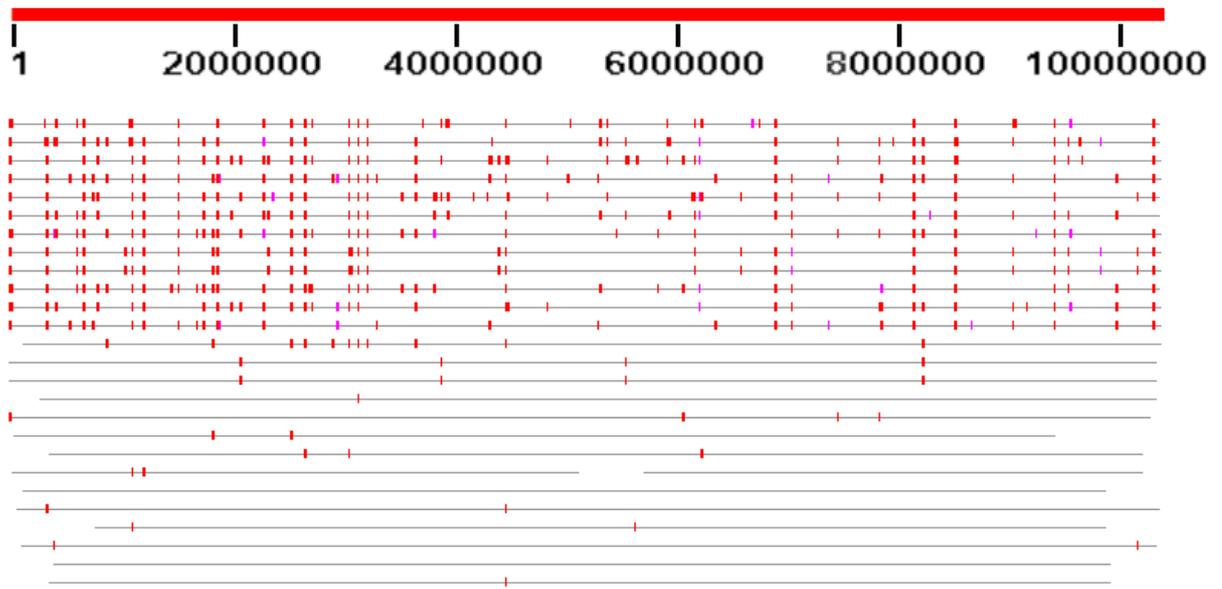


Figure 3.1: BLAST results showing poor homology between claimed *Streptomyces species* that varies from <1% to >99% indicating poor taxonomy of the current Genus *Streptomyces*.

Accession	Description	Query coverage	Max ider
GU384063.1	Streptomyces griseus strain DSM 40931 Fe-containing superoxide dismutase (sodF)	100%	100%
GU384062.1	Streptomyces limosus strain CB218 Fe-containing superoxide dismutase (sodF)	100%	99%
GU384027.1	Streptomyces limosus strain CB143 Fe-containing superoxide dismutase (sodF)	100%	99%
GU384025.1	Streptomyces limosus strain CB141 Fe-containing superoxide dismutase (sodF)	100%	99%
GU384060.1	Streptomyces limosus strain E961 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384059.1	Streptomyces limosus strain E953 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384050.1	Streptomyces limosus strain E956 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384049.1	Streptomyces limosus strain E948 Fe-containing superoxide dismutase (sodF)	100%	98%
GU383967.1	Streptomyces limosus strain 651 Fe-containing superoxide dismutase (sodF)	100%	98%
GU383966.1	Streptomyces limosus strain DSM 40131 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384045.1	Streptomyces limosus strain CB171 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384036.1	Streptomyces limosus strain CB152 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384035.1	Streptomyces limosus strain CB151 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384034.1	Streptomyces limosus strain CB150 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384033.1	Streptomyces limosus strain CB149 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384032.1	Streptomyces limosus strain CB148 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384030.1	Streptomyces limosus strain CB146 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384028.1	Streptomyces limosus strain CB144 Fe-containing superoxide dismutase (sodF)	100%	98%
GU384057.1	Streptomyces griseus strain CB206 Fe-containing superoxide dismutase (sodF)	100%	97%
GU384066.1	Streptomyces cinnamoneus strain DSM 40005 Fe-containing superoxide dismutase (sodF)	98%	91%
GU383996.1	Streptomyces glaucescens strain DSM 40155 Fe-containing superoxide dismutase (sodF)	100%	90%
GU383995.1	Streptomyces violaceoruber strain DSM 40049 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383994.1	Streptomyces coelicolor strain NRRL B-12000 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383993.1	Streptomyces sp. BTG 4-738 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383992.1	Streptomyces sp. BTG 4-723 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383991.1	Streptomyces sp. BTG 723 I Fe-containing superoxide dismutase (sodF)	98%	90%
GU383990.1	Streptomyces coelicolor strain M110 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383989.1	Streptomyces sp. BTG 717 I Fe-containing superoxide dismutase (sodF)	98%	90%
GU383988.1	Streptomyces sp. BTG 6-708 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383987.1	Streptomyces sp. BTG 6-715 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383986.1	Streptomyces sp. BTG 4-759 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383985.1	Streptomyces sp. BTG 4-758 Fe-containing superoxide dismutase (sodF)	98%	90%
GU383984.1	Streptomyces lividans strain 1326 Fe-containing superoxide dismutase (sodF)	98%	90%
GU384024.1	Streptomyces roseochromogenes strain DSM 40463 Fe-containing superoxide dismutase (sodF)	100%	88%
GU384013.1	Streptomyces griseobrunneus strain DSM 40066 Fe-containing superoxide dismutase (sodF)	100%	88%
GU384013.1	Streptomyces bacillaris strain DSM 40598 Fe-containing superoxide dismutase (sodF)	100%	87%
AL939107.1	Streptomyces coelicolor A3(2) complete genome; segment 4/29	98%	88%
AF099015.1	Streptomyces coelicolor strain A3(2) integrase (int), Fe-containing superoxide dismutase (sodF)	98%	88%
GU384061.1	Streptomyces sp. E1002 Fe-containing superoxide dismutase (sodF) gene, partial	100%	87%
GU384056.1	Streptomyces griseus strain DSM 40878 Fe-containing superoxide dismutase (sodF)	100%	87%
GU384055.1	Streptomyces griseus strain DSM 40759 Fe-containing superoxide dismutase (sodF)	100%	87%
GU384054.1	Streptomyces griseus strain DSM 40670 Fe-containing superoxide dismutase (sodF)	100%	87%
GU384048.1	Streptomyces griseus strain DSM 40658 Fe-containing superoxide dismutase (sodF)	100%	87%

Figure 3.2: Superoxide Dismutase (SOD) BLAST results showing major variations among different *Streptomyces* spp. yet consistent within each species. Some inconsistencies suggest that such species may have been misidentified; red arrows.

3.2 Specificity of the primer:

The selected primers were specific for *Streptomyces*, they did not react with DNA extracted from non-*Streptomyces* bacteria, a collection of known and unknown Gram positive and Gram negative bacterial DNA that was extracted as described in the methods section. No PCR amplification was seen as judged by agarose-ethidium bromide gel electrophoresis, this result was confirmed upon repetition on a different mixture of bacteria on a different day.

The selected primers of the different *Streptomyces* loci were all functional; each primer pair produced the expected amplicon (Table 2.1); Some isolates were positive producing amplifications with certain primer pairs while negative (no amplification) with others indicating the suitability of these loci to discriminate between species in accordance with the hypothesis. SOD was positive for 26 isolates, σ f was positive for 14 isolates, PPD was positive for 3 isolates, TR was positive for 2 isolates, aaa was positive for 4 isolates and IS was positive for 4 isolates.

3.3 Screening of selected isolates with all primer pairs applied as multiplexes for clustering:

Table 2.2 shows the different primer mixtures, 2mix and 4mix; amplicon sizes are indicated as well. When these mixes along with 16S primers were applied as four separate PCR reactions per *Streptomyces* isolate, the results (Figure 3.4, 3.5 and 3.6) showed different PCR patterns for several isolates.

3.4 16S rDNA groups (A, B, and C):

Figure 3.3A shows four of the isolates; QUBC 259, 267, 291 and 311. Additional results are shown in figures 3.3B and C. The discriminatory ability of 16S primers to separate isolates into three (16s-A, B and C) groups was demonstrated, with isolates belonging to the A, B or C ribosomal groups.

Isolates that produced negative results (16S-D) were excluded from this study.

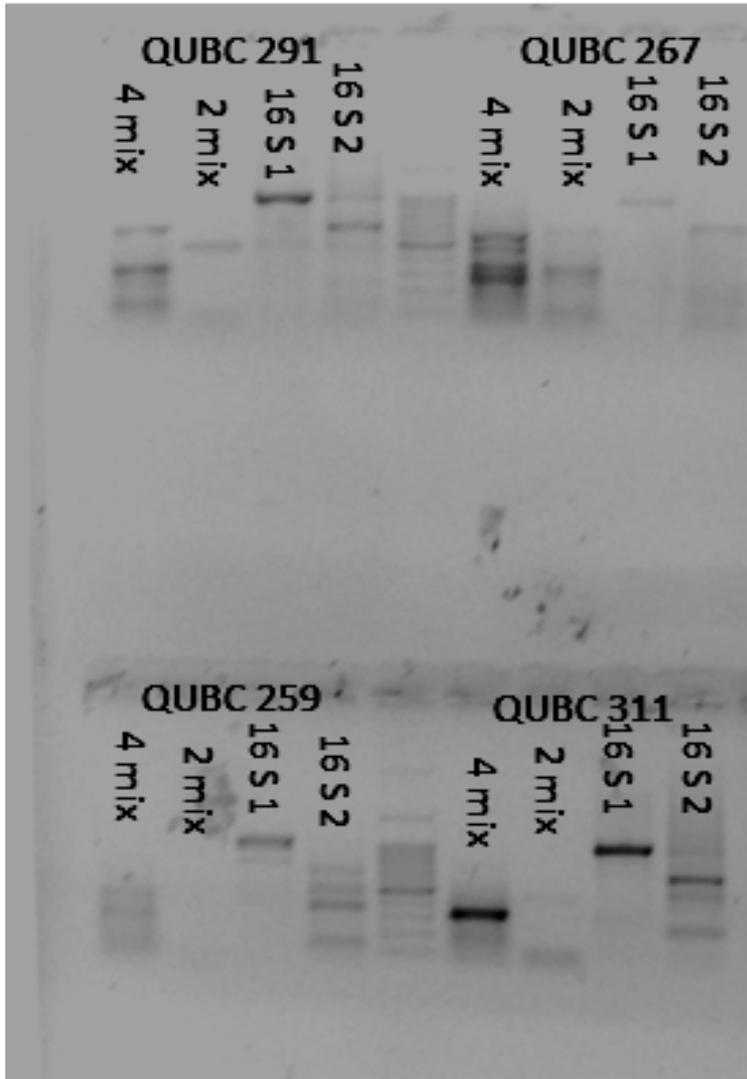


Figure 3.3 A: Three isolates belong to 16s-A group, QUBC 259 belongs to 16S-B since only one band is showing in the first reaction of 16s reaction.

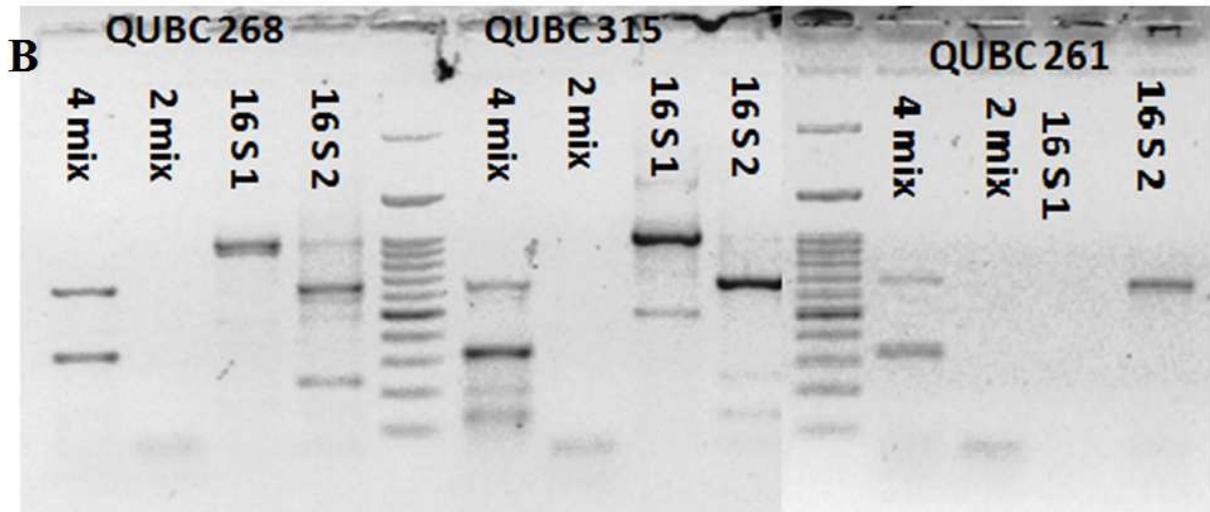


Figure 3.3 B: Three isolates Showing QUBC268 (16S-B); QUBC 315 (16S-A), and QUBC 261 (16s-C) yet all three isolates share the same profile (2mix) with no bands.

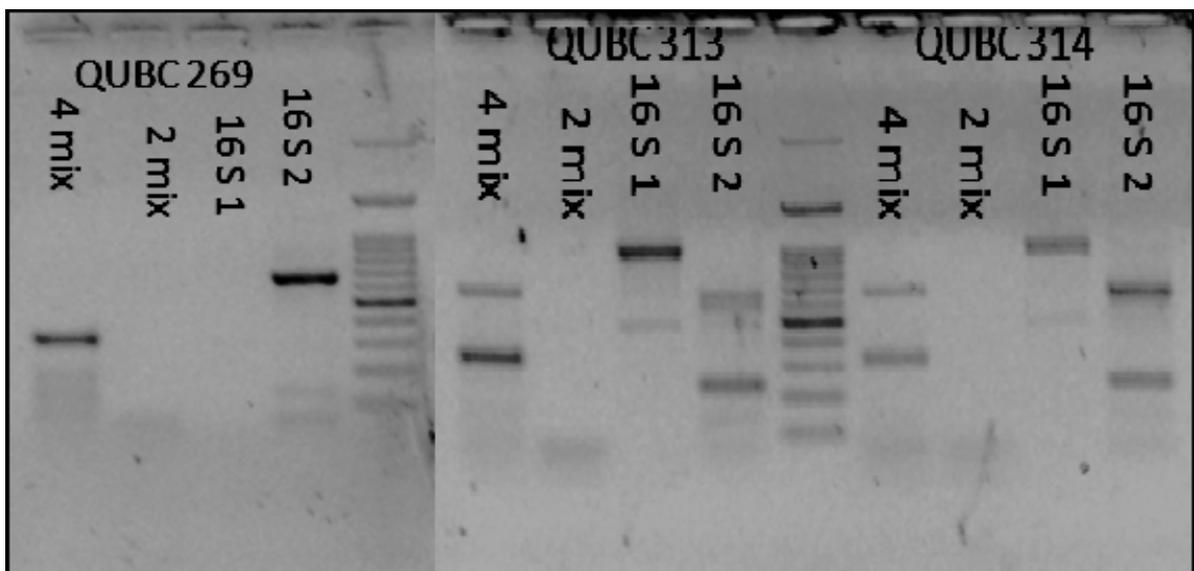


Figure 3.3 C: Three additional isolate showing QUBC 269 (16s-C), and QUBC 313 and 314 (16S-B). All tested isolates produced the 333bp band (Figure 3.3; Table 3.1) indicative of superoxide dismutase (4mix; smaller band of PCR multiplex reaction). However, sigma factor locus (indicated by the 709 bp in 4 mix reaction) was negative in isolates QUBC 259, 269, and 311. Results for all tested isolate are summarized in Table 3.1. The presence of a locus (positive ;+) or its absence (negative ;-).

3.5 Other loci:

Isolates QUBC 96, 184, 202, and 286 were positive for amino acid adenylation domain protein (aaa) representing 14% of the 28 tested isolates. Pyruvate Phosphate Dikinase locus was identified in the three isolates QUBC 267, 285, and 286 (10.7%), only one isolate (QUBC 286) that contained both (aaa and PPD) loci which makes it unique among the 28 tested isolates.

3.6 Assignment of isolates to a specific Cluster:

Table 3.1: Summary of all tested isolates showing the presence or absence of specified locus.

QUBC	Genes			SOD	σf	PPD	TR	aaa	IS	Generated Cluster #
	Fw1	Fw2.R	16S rDNA							
	1+,+ A	2+,+ B	-,+ C							
12	-	-	+	+	-	-	-	-	-	160
269	-	-	+	+	-	-	-	-	-	160
96	-	-	+	+	+	-	-	+	-	142
184	+	-	-	+	-	-	-	+	-	30
267	+	-	-	+	+	+	-	-	+	7
291	+	-	-	+	-	-	-	-	+	31
210	-	-	+	+	-	-	-	-	+	159
285	-	-	+	+	-	+	+	-	-	148
286	-	+	-	+	+	+	-	+	+	69
202	-	-	+	+	-	-	+	+	-	154
289	+	-	-	+	+	-	-	-	-	16
290	+	-	-	+	+	-	-	-	-	16
315	+	-	-	+	+	-	-	-	-	16
259	-	+	-	+	-	-	-	-	-	96
102	-	+	-	+	-	-	-	-	-	96
103	-	+	-	+	-	-	-	-	-	96
261	-	-	+	+	+	-	-	-	-	144
288	-	-	+	+	+	-	-	-	-	144
157	-	-	+	+	+	-	-	-	-	144
273	-	-	+	+	+	-	-	-	-	144
313	-	+	-	+	+	-	-	-	-	80
314	-	+	-	+	+	-	-	-	-	80
287	-	+	-	+	+	-	-	-	-	80
268	-	+	-	+	+	-	-	-	-	80
101	+	-	-	+	-	-	-	-	-	32
311	+	-	-	+	-	-	-	-	-	32

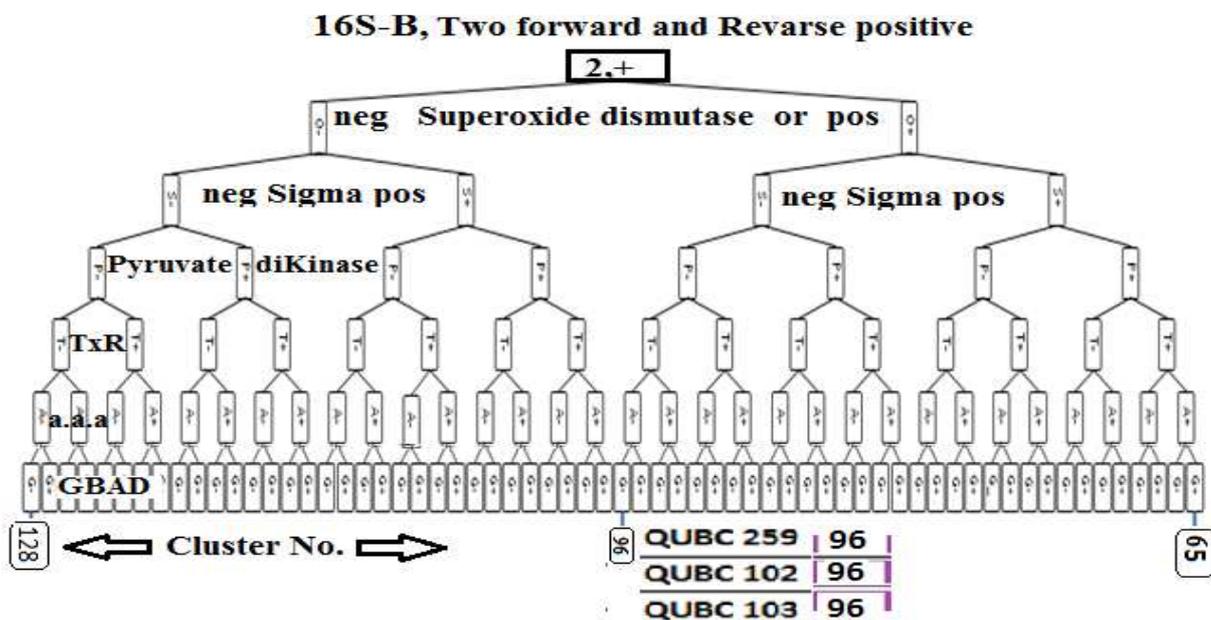


Figure 3.5: Flowchart of the 16S-B group Isolates belong to this group include QUBC 286 , in cluster 69. Isolates QUBC102, 103 and 259 belong to cluster 96. Isolates QUBC 313, 314, 287 and 268 belong to cluster 80.

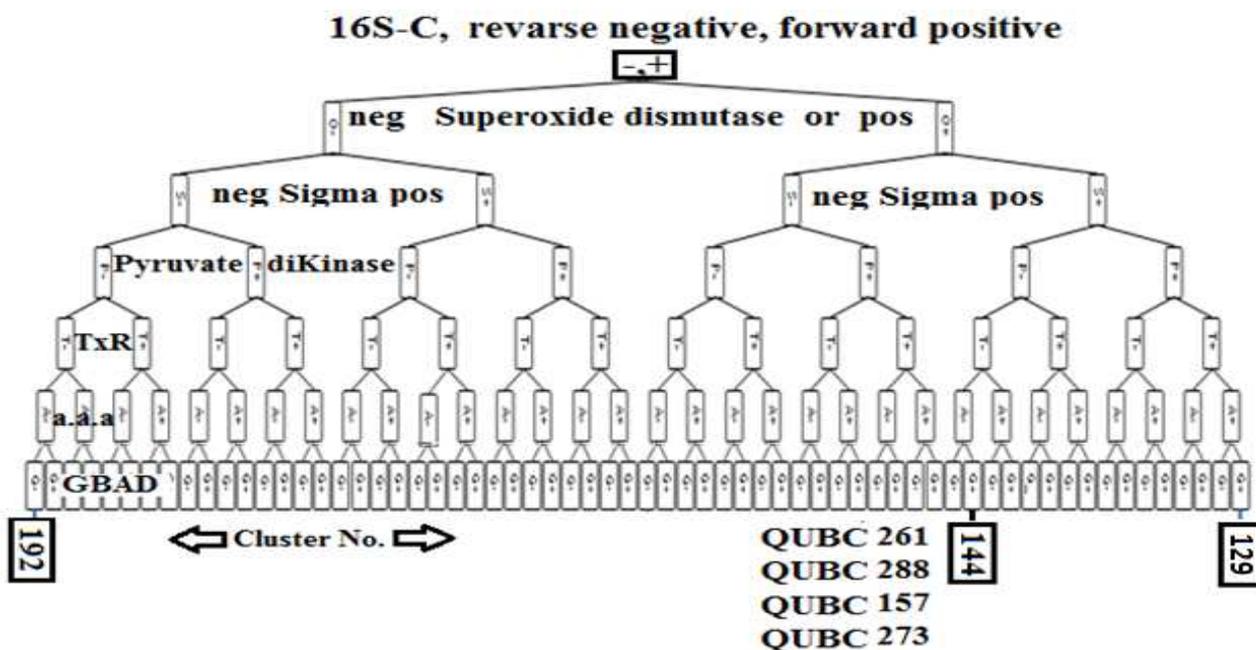


Figure 3.6 : Flowchart of the 16S-C group Isolates belong to this group include QUBC 261, 288, 157, and 273 in cluster 144. Isolates QUBC 12 and 269 belong to cluster 160. QUBC 96 in cluster 142. QUBC 210 in cluster 159. QUBC 285 in cluster 148. QUBC 202 in cluster 154.

3.8 Superoxide dismutase amplicon sequencing and generation of a phylogenic tree:

Since most isolates have shown presence of Superoxide dismutase amplicons, this locus was selected for sequencing in order to gain some idea about the relationships of clusters to actual DNA sequences. Figure 3.7 shows the obtained phylogenic tree. DNA sequencing, and construction of a phylogenic tree based on multiple alignment; Bioedit software (<http://www.mbio.ncsu.edu/BioEdit/page2.html>) (53).

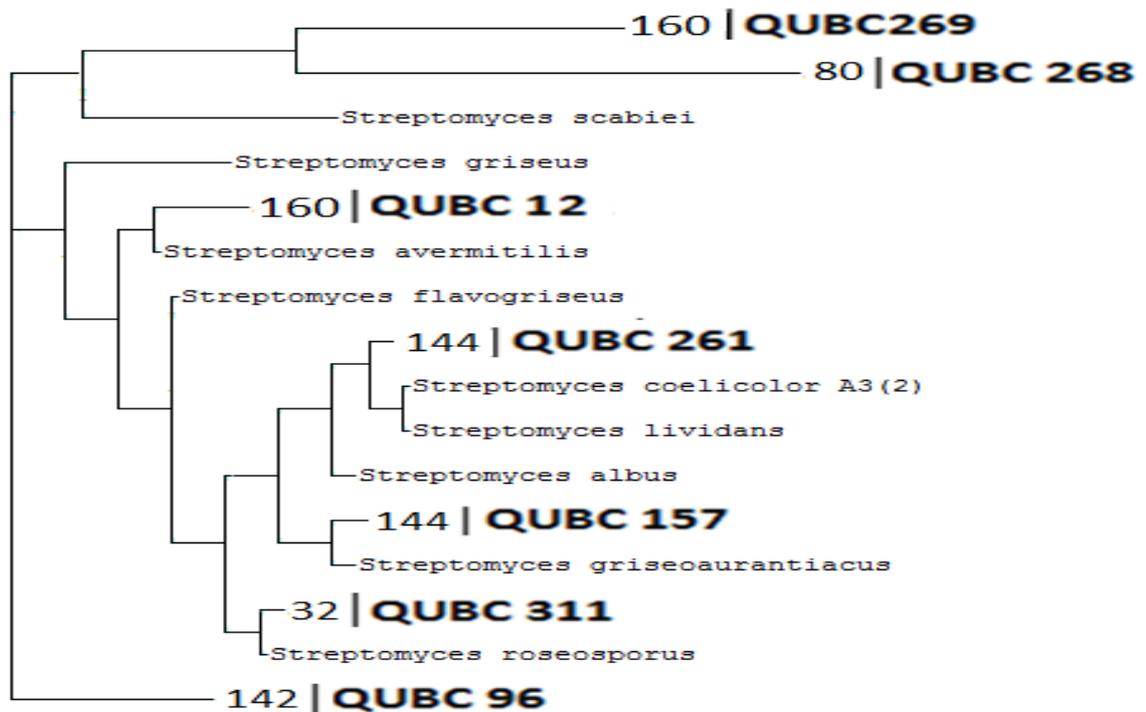


Figure 3.7: Superoxide dismutase amplicon (333bp) was sequenced (QUBC isolates, with their cluster number indicated, additional published (SOD) sequences (scientific binomial names) were incorporated to construct this phylogenic tree. Several observations can be made based on this tree (see discussion).

3.9 One soil sample:

Five putative isolates were obtained from the same CCG agar plate of a soil sample (Germany). The five isolates were used to challenge the multilocus ability to resolve different species and to identify similar or identical species. The results are shown in table 3.2 and figure 3.8

Table 3.2: Five putative isolates were obtained from the same soil sample

Isolate	Cluster	16S group
QUBC 286	69	B
QUBC 287	80	B
QUBC 288	144	C
QUBC 289	16	A
QUBC 290	16	A



Figure 3.8 : Two or three different morphological isolates from one soil sample (Germany) on the CCG medium. Isolate QUBC 287-290 were analyzed and showed two similar isolates and four different clusters (Table 3.3)

3.10 Distribution of tested isolates among hypothesized multiplex clusters:

Table 3.3 shows that the 26 bacterial isolate were distributed among 14 different clusters, the soil samples were from different locations, four isolates sharing cluster 144 all came from different countries, QUBC 157 from Palestine/Ramallah; QUBC 261 from Qatar, QUBC 288 from Germany, and QUBC 273 from Lebanon. Others that share the same cluster 80 came from the same soil sample (Kubar, Palestine), two German isolates shared the same cluster 16. Rare isolates came from unique environments; desert sand showed unique clusters 30; Cyprus sand (QUBC 291) showed cluster 31; Sharjeh isolate QUBC 96 is from the Arabian Gulf and was previously identified as an oceanic *S. krainski*.

Table 3.3: Assigned isolates to the hypothesized generated clusters

Cluster	Isolate	Source	Cluster	Isolate	Source
160	QUBC 12	ابوديس	96	QUBC 259	قطر 8
160	QUBC 269	الجرف/النقب	96	QUBC 102	ابوديس
142	QUBC 96	الشارقة ² <i>S.krainski</i>	96	QUBC 103	ابوديس
30	QUBC 184	رمل صحراوي	144	QUBC 157	ابو خليل رصاصي
7	QUBC 267	بيت حانون	144	QUBC 261	قطر 10 بني
31	QUBC 291	1X قبرص	144	QUBC 288	المانيا
159	QUBC 210	حديقتي شوكلاته	144	QUBC 273	بيروت/ضبية
148	QUBC 285	تونس س.هـ	80	QUBC 313	كوبر بيضاء شعاعيه
69	QUBC 286	المانيا	80	QUBC 314	كوبر بيضاء
154	QUBC 202	رام الله/المختبر	80	QUBC 287	المانيا
16	QUBC 289	المانيا	80	QUBC 268	الجرف/النقب
16	QUBC 290	المانيا	32	QUBC 101	ابوديس
16	QUBC 315	كوبر رمادي	32	QUBC 311	Aالسويد

Chapter 4

Discussion and Conclusions

Genotypic analysis has contributed greatly to our knowledge about *Streptomyces* taxonomy, as gene sequences are normally not affected by growth conditions, medium composition, or environmental factors. Genotypic instability has been recognized that may affect genotypic classification as well (9). This may explain the inconsistency observed when BLAST is used to align a single species against itself (Figure 3.1 and 3.2). On the contrary, fluctuation of such factors is known to affect reproducibility of phenotypic features used as taxonomic traits.

The sequence analysis and multiple alignments and the construction of a phylogenetic tree showed that most isolates were different and can be discriminated into one of 192 possible clusters based on a single PCR run. Clustering was made simple; it was composed of four reactions (Two 16S mixtures and two mixtures of 6 loci 4mix and 2mix) run in 4 lanes on 1.5% agarose gel. The results showed excellent resolution of all tested *Streptomyces spp.* suggesting that the multilocus clustering approach will be able to solve several important issues: 1- Breakdown a complex Taxa down to small groups based on 16S rDNA (A, B, and C). 2- The system allowed the use of the same trait or character three times, one time per group (e.g. a SOD positive or negative could only be used once if all members were in one group, but can be applied three times in the ABC 3-group-arrangement. In other words, the system has introduced a 3-folds reduction in complexity. 4- Because 16S was utilized to breakdown the complex relying on the presence or absence of a semi-conservative primer sequences (Table 2.1), positive results meant *Streptomyces spp.*

This study provided the necessary tables and figures needed to allow unambiguous clustering of current >570 *Streptomyces spp.* Mathematically, each cluster should contain an average of 3 species per cluster; the range may be wide with some clusters showing more species. Alternatively, some clusters may contain one or two species only. In case of increased numbers of species double or triple (e.g ~1500); each cluster will be able to accommodate nine species on average, such species can be distinguished based on their stable characteristics, or additional multiplex that contains two or three additional loci. Therefore, the method can be adapted to distinguish all isolates down to species level. It must be emphasized that clustering

is a process meant to reduce the number of species that need to be discriminated from each other. Such a method will alleviate the need to direct classification of any given *Streptomyces* isolate; a process that requires special training and extensive expertise in the *Streptomyces* taxonomy. It must be clarified that clustering does not necessitate that species within a cluster be phylogenetically homogeneous, instead the more diverse they are, the easier it is to distinguish them from each other.

In this sense, multiplex clustering substitutes phenotypic traits with molecular traits that are much more stable, reliable; flexible, and amenable to analysis; they are unlikely to be misinterpreted as it is the case of color-dependent reactions or time dependent-reactions and ambiguities associated with phenotypic methods or 16S ribotyping alone. An added advantage, sequencing of a selected locus such as Super Oxide Dismutase may identify the species and confirm its identity.

The results showed the ability of the system to discriminate between isolates, while indicating a possible replication of isolates QUBC 289 and 290 (Table 3.3, Figure 3.7, and Figure 3.8). Additional testing, morphological features, or amplicon sequencing may be used to determine the relationship between these two isolates. The German soil sample produced five different isolates (Table 3.2).

As presented in section 3.7, the multiplex clustering was able to discriminate between isolates obtained from the same soil sample, yet showed its capacity to identify similar isolates obtained from unrelated samples or the same soil sample. Samples from different environments showed isolates in different clusters. The results were consistent, valid, and reliable. The results support the hypothesis that multilocus clustering can discriminate between different isolates while identifying similar isolates as such. The 26 tested isolates were distributed among 14 different clusters, resolution ($R=26/14= 1.85$ species per cluster) was also consistent with the predicted resolution (3 species per cluster) that was theoretically calculated ($570/192= 2.97$ sp./cluster).

The dendrogram (figure 3.7) shows that QUBC 311 of cluster 32 appears to be closely related to *S. roseosporus* based on SOD sequence. Accordingly, incorporation of an unknown species SOD sequence with published SOD sequences may allow species identification and cluster

assignment to some extent to known published sequences, it may be reasonable to assume that *S. roseosporus* may be closely related to cluster 32. Similarly QUBC 311 of cluster 32 may be related (similar or identical) to *S. roseosporus*. The problem is that current taxonomy of *Streptomyces* spp. is not reliable according to BLAST results (section 3.1).

When an isolate produced all negative PCR results, DNA was shown to be of a different bacterium using the G7 primer multiplex for bacterial detection according to the described Universal Method (Barghouthi 2011) (49).

Recommendation:

Finally, after this study, it is recommended that all ATCC of *Streptomyces* be provided from the research center and run them through this designed system to classify all *Streptomyces* spp. into generated taxonomic clusters, to know exactly how they will cluster and if the system needs tuning to do. By clustering *Streptomyces* spp. it will be easy to identify any unknown isolates and to what cluster they belong, and further more it is possible to limit an unknown isolate to a particular cluster. Future efforts should be directed toward what was mentioned above, and to additional studies of each cluster to add specific tests or characters to that could aid in species and strain identification.

References:

1. Lezhava, A., T. Mizukami, T. Kajitani, D. Kameoka, M. Redenbach, H. Shinkawa, O. Nimi and H. Kinashi, Physical map of the linear chromosome of *Streptomyces griseus*. Journal of bacteriology, 1995. **177**(22): p. 6492-8.
2. Goshi, K., T. Uchida, A. Lezhava, M. Yamasaki, K. Hiratsu, H. Shinkawa and H. Kinashi, Cloning and analysis of the telomere and terminal inverted repeat of the linear chromosome of *Streptomyces griseus*. Journal of bacteriology, 2002. **184**(12): p. 3411-5.
3. Flardh, K. and M.J. Buttner, *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. Nature reviews. Microbiology, 2009. **7**(1): p. 36-49.
4. Loria, R., J. Kers, and M. Joshi, Evolution of plant pathogenicity in *Streptomyces*. Annual review of phytopathology, 2006. **44**: p. 469-87.
5. Dunne, E.F., W.J. Burman, and M.L. Wilson, *Streptomyces* pneumonia in a patient with human immunodeficiency virus infection: case report and review of the literature on invasive *Streptomyces* infections. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 1998. **27**(1): p. 93-6.
6. Develoux, M., J. Audoin, J. Treguer, J. M. Vetter, A. Warter and A. Cenac, Mycetoma in the Republic of Niger: clinical features and epidemiology. The American journal of tropical medicine and hygiene, 1988. **38**(2): p. 386-90.
7. Chavez, G., R. Estrada, and A. Bonifaz, Perianal actinomycetoma experience of 20 cases. International journal of dermatology, 2002. **41**(8): p. 491-3.
8. Rose, C.E., 3rd, J.M. Brown, and J.F. Fisher, Brain abscess caused by *Streptomyces* infection following penetration trauma: case report and results of susceptibility analysis of 92 isolates of *Streptomyces* species submitted to the CDC from 2000 to 2004. Journal of clinical microbiology, 2008. **46**(2): p. 821-3.
9. Anderson, A.S. and E.M. Wellington, The taxonomy of *Streptomyces* and related genera. International journal of systematic and evolutionary microbiology, 2001. **51**(Pt 3): p. 797-814.
10. Berd, D., Laboratory identification of clinically important aerobic actinomycetes. Applied microbiology, 1973. **25**(4): p. 665-81.
11. Singh, V., V. Praveen, F. Khan and C. K. Tripathi, Phylogenetics of an antibiotic producing *Streptomyces* strain isolated from soil. Bioinformation, 2009. **4**(2): p. 53-8.
12. Gehring, A.M., N.J. Yoo, and R. Losick, RNA polymerase sigma factor that blocks morphological differentiation by *Streptomyces coelicolor*. Journal of bacteriology, 2001. **183**(20): p. 5991-6.

13. Miller, P.H., L.S. Wiggs, and J.M. Miller, Evaluation of API An-IDENT and RapID ANA II systems for identification of Actinomyces species from clinical specimens. Journal of clinical microbiology, 1995. **33**(2): p. 329-30.
14. Guo, Y., W. Zheng, X. Rong and Y. Huang, A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. International journal of systematic and evolutionary microbiology, 2008. **58**(Pt 1): p. 149-59.
15. Staley, J.T., The Phylogenomic Species Concept for Bacteria and Archaea. Microbe, 2009. **4**: p. 361-365.
16. Waksman, S.A. and A.T. Henrici, The Nomenclature and Classification of the Actinomycetes. Journal of bacteriology, 1943. **46**(4): p. 337-41.
17. Waksman, S.A., On the Classification of Actinomycetes. Journal of bacteriology, 1940. **39**(5): p. 549-58.
18. Minnikin, D.E. and A.G. O'Donnell, Actinomycete envelope lipid and peptidoglycan composition., in The Biology of the Actinomycetes, M. Goodfellow, M. M.Mordarski, and S.T. Williams, Editors. 1984, Academic Press: London. p. 337-388.
19. Bronz, I., I. Olsen, M. Haapasalo and A. Van Winkelhoff, Multivariate analyses of fatty acid data from whole-cell methanolysates of Prevotella, Bacteroides and Porphyromonas spp. Journal of General Microbiology, 1991. **137**(6): p. 1445-1452.
20. Gudmestad, N.C., P.J. Henningson, and W.M. Bugbee, Cellular fatty acid comparison of strains of Corynebacterium michiganense subsp. sepedonicum from potato and sugar beet. Canadian Journal of Microbiology, 1988. **34**(6): p. 716-722.
21. Saddler, G.S., A. G. O'Donnell, M. Goodfellow and D. E. Minnikin, SIMCA Pattern Recognition in the Analysis of Streptomycete Fatty Acids. Journal of General Microbiology, 1987. **133**(5): p. 1137-1147.
22. Sanglier, J.-J., D. Whitehead, G. S. Saddler, E. V. Ferguson and M. Goodfellow, Pyrolysis mass spectrometry as a method for the classification, identification and selection of actinomycetes. Gene, 1992. **115**(1-2): p. 235-242.
23. Goodfellow, M., C. Lonsdale, A.L. James and O.C. MacNamara, Rapid biochemical tests for the characterisation of streptomycetes. FEMS Microbiology Letters, 1987. **43**(1): p. 39-44.
24. Korn-Wendisch, F. and H.J. Kutzner, The family Streptomycetaceae, in The Prokaryotes, A. Balows, et al., Editors. 1992, Springer: New York. p. 921-995.
25. Manchester, L., B. Pot, K. Kersters and M. Goodfellow, Classification of *Streptomyces* and *Streptoverticillium* Species by Numerical Analysis of Electrophoretic Protein Patterns. Systematic and Applied Microbiology, 1990. **13**(4): p. 333-337.

26. Paradis, E., C. Goyer, N. Hodge, R. Hogue, R. Stall and C. Beaulieu, Fatty Acid and Protein Profiles of *Streptomyces* scabies Strains Isolated in Eastern Canada. *International Journal of Systematic Bacteriology*, 1994. **44**(3): p. 561-564.
27. Oh, C.-h., M.-h. Ahn, and J.-h. Kim, Use of electrophoretic enzyme patterns for streptomycete systematics. *FEMS Microbiology Letters*, 1996. **140**(1): p. 9-13.
28. Becker, B., M. P Lechevalier, R. E. Gordon and H. A. Lechevalier, Rapid Differentiation between *Nocardia* and *Streptomyces* by Paper Chromatography of Whole-Cell Hydrolysates. *Applied microbiology*, 1964. **12**: p. 421-3.
29. Ridell, M., G. Wallerström, and S.T. Williams, Immunodiffusion Analyses of Phenetically Defined Strains of *Streptomyces*, *Streptoverticillium* and *Nocardiosis*. *Systematic and Applied Microbiology*, 1986. **8**(1-2): p. 24-27.
30. Kirby, R. and E.P. Rybicki, Enzyme-linked immunosorbent assay (ELISA) as a means of taxonomic analysis of *Streptomyces* and related organisms. *Journal of general microbiology*, 1986. **132**(7): p. 1891-4.
31. Wipat, A., E.M.H. Wellington, and V.A. Saunders, Monoclonal antibodies for *Streptomyces lividans* and their use for immunomagnetic capture of spores from soil. *Microbiology*, 1994. **140**(8): p. 2067-2076.
32. Williams, S.T., M. Goodfellow, G. Alderson, E. M. Wellington, P. H. Sneath and M. J. Sackin, Numerical classification of *Streptomyces* and related genera. *Journal of general microbiology*, 1983. **129**(6): p. 1743-813.
33. Williams, S.T., M. Goodfellow, E. M. Wellington, J. C. Vickers, G. Alderson, P. H. Sneath, M. J. Sackin and A. M. Mortimer, A probability matrix for identification of some Streptomycetes. *Journal of general microbiology*, 1983. **129**(6): p. 1815-30.
34. Willcox, W.R., S. P. Lapage, S. Bascomb and M. A. Curtis, Identification of bacteria by computer: theory and programming. *Journal of general microbiology*, 1973. **77**(2): p. 317-30.
35. Ismail, A., Numerical Assessment of Mycelium Color in Classification of Some *Streptomyces* Isolates. *International Journal Of Agriculture & Biology*, 2006. **8**(6): p. 872-875.
36. Witt, D. and E. Stackebrandt, Unification of the Genera *Streptoverticillum* and *Streptomyces*, and Amendment of *Streptomyces* Waksman and Henrici 1943, 339AL. *Systematic and Applied Microbiology*, 1990. **13**(4): p. 361-371.
37. BEYAZOVA, M. and M.P. LECHEVALIER, Taxonomic Utility of Restriction Endonuclease Fingerprinting of Large DNA Fragments from *Streptomyces* Strains. *International Journal of Systematic Bacteriology*, 1993. **43**(4): p. 674-682.

38. Anzai, Y., T. Okuda, and J. Watanabe, Application of the random amplified polymorphic DNA using the polymerase chain reaction for efficient elimination of duplicate strains in microbial screening. II. Actinomycetes. *The Journal of antibiotics*, 1994. **47**(2): p. 183-93.
39. Mehling, A., U.F. Wehmeier, and W. Piepersberg, Application of random amplified polymorphic DNA (RAPD) assays in identifying conserved regions of actinomycete genomes. *FEMS Microbiology Letters*, 1995. **128**(2): p. 119-25.
40. Stackebrandt, E., W. Liesack, and D. Witt, Ribosomal RNA and rDNA sequence analyses. *Gene*, 1992. **115**(1-2): p. 255-60.
41. Stackebrandt, E., D. Witt, C. Kemmerling, R. Kroppenstedt and W. Liesack, Designation of Streptomyces 16S and 23S rRNA-based target regions for oligonucleotide probes. *Applied and environmental microbiology*, 1991. **57**(5): p. 1468-77.
42. Ludwig, W. and K.H. Schleifer, Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS microbiology reviews*, 1994. **15**(2-3): p. 155-73.
43. Huddleston, A.S., N. Cresswell, M. C. Neves, J. E. Beringer, S. Baumberg, D. I. Thomas and E. M. Wellington, Molecular detection of streptomycin-producing streptomycetes in Brazilian soils. *Applied and environmental microbiology*, 1997. **63**(4): p. 1288-97.
44. Liu, K., S.J. Knabel, and E.G. Dudley, rhs genes are potential markers for multilocus sequence typing of Escherichia coli O157:H7 strains. *Applied and environmental microbiology*, 2009. **75**(18): p. 5853-62.
45. Barghouthi, S., Ayesha M, A semi selective medium for the isolation and identification of *Streptomyces* spp. from the soil samples. Abstracted submitted to The Sixth Palestinian Conference of laboratory medicine 12-13 March 2010, Al-Quds University.
46. Genomes. 2011; Available from: <http://www.ncbi.nlm.nih.gov/genomes/genlist.cgi?taxid=2&type=0&name=Complete%20Bacteria>.
47. Table. 2011; Available from: http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi.
48. BioMath Calculators. 2011; Available from: http://www.promeqa.com/techserv/tools/biomath/calc11.htm#melt_results.
49. Barghouthi, S., A Universal Method for the Identification of Bacteria Based on General PCR Primers. *Indian Journal of Microbiology*, 2011. **DO: 10.1007/s12088-011-0122-5**.
50. Barghouthi, S. and D. Al Zughayyar, Detection of Neisseria meningitidis and unknown Gammaproteobacteria in cerebrospinal fluid using the two-step universal method. *African Journal of Microbiology Research*, 2012. **6**(14): p. 3415-3424.

51. Kieser T, B.M., Buttiner MJ, Chater KF, Hopwood DA, Practical *Streptomyces* Genetics2000: John Innes Foundation. 613.

52. Don, R.H., P. T. Cox, B. J. Wainwright, K. Baker and J. S. Mattick, 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic acids research, 1991. **19**(14): p. 4008.

53. Hall, T. (2005): Bioedit. Ibis Biosciences, Carlsbad, CA 92008.

54. Baylis, H.A. and M.J. Bibb, The nucleotide sequence of a 16S rRNA gene from *Streptomyces coelicolor* A3(2). Nucleic acids research, 1987. **15**(17): p. 7176.

Appendix:

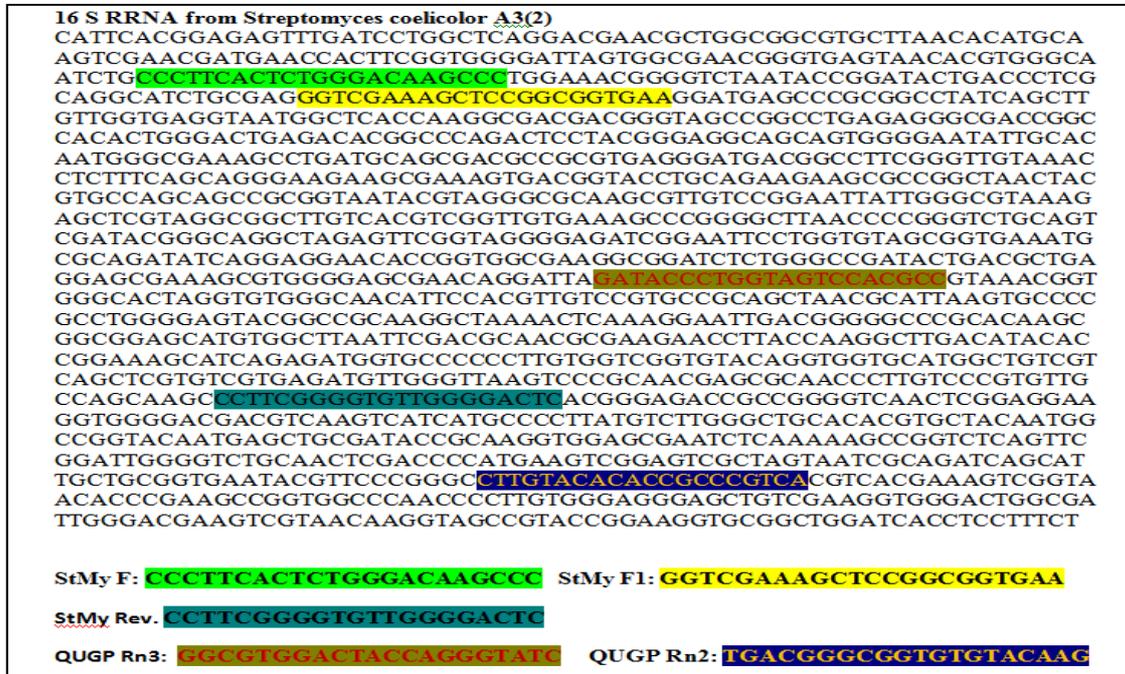


Figure A1: *Streptomyces coelicolor* 16S rRNA and locations of all 16S primers see reference (54).

The figure show sequence of 16S RRNA and the highlighted sequence demonstrate five primers which were used as two set of multiplex (Table 2.1 and 2.2). The first set composed of two forward primers with one reverse primer, the other one composed of same set in additional to other two reverse primers. Possible PCR amplicons of 16S as applied for *Streptomyces* are shown in Table 2, six possible amplicons for each isolate are indicated.

Table A1: Size of the PCR product in the two sets of 16S RRNA

Primers	StMy R	QUGP Rn3	QUGP Rn2
StMy F	1005bp	659	1265
StMy F 2	934 bp	752	1172

Figures A2: Loci used in this study showed discriminatory power among published genomic sequences of *S. species* (see section 2.2).

Superoxide dismutase from *Streptomyces coelicolor* A3(2)

TCACGGCTTGAGGAGCAGGCTGTCGCCGCGCTCCTTGCGGGCGGCGTGACGCTTGGCCACGTCCTGC
CAGTTGACGACGGCCACATCGCCTCGATGAAGTCGACCTTCTGGTTCTTGTACTGCAGGTAGAAGG
CGTGCTCCAGGCGTGAAGACGAGGATCGGGGTGGCGCCCTGGCCGACGTTGCCCTGGTGGTCGTA
GACCTGCTCGACGATCAGCCGGCCGCTGAGCGGCTCGTAGGCGAGGACGCCCCAGCCGGAGCCCTGG
GTGGTGGCGGCGGCCTTGGTCAGCTGGGCCTTGAAGCCGGCGAAGGAACCGAAGGACTCGGTGATC
GCGTCGGCCAGGTCACCGACACCGTCCTGGGCGAGGGGCTCGCCGCCGCGTCCTTCGGGCCCCTCA
TGTTCTGCCAGTAGATCGAGTGCAGAATGTGCCCGGAGAGGTGGAACGCGAGGTTCTTCTCCAGGCC
GTTGATCGAACCCACGACTCCTTCTCCCGCGCCTCGGCCAGCTGCTCCAGGGTGTCGTTTCGACCCT
TCACGTACGCCGCATGATGCTTGTCTGTTGTCAGCTCGATGATCTCCGGGCTGATGACCGGAGCCAG
CGCCGAGTAGTCGTACGGAAGTTCCGGAAGCGTGTAAGCGCCAT

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AL939107.1	Streptomyces coelicolor A3(2) complete genome; segment 4/29	1197	1197	100%	0.0	100%	
AF099015.1	Streptomyces coelicolor strain A3(2) integrase (int), Fe-containing superox	1197	1197	100%	0.0	100%	G
AP009493.1	Streptomyces griseus subsp. griseus NBRC 13350 DNA, complete genome	784	784	94%	0.0	90%	
AF141866.2	Streptomyces griseus Fe-Zn-superoxide dismutase SodF (sodF), transcript	784	784	94%	0.0	90%	
GU384068.1	Streptomyces galbus strain DSM 40480 Fe-containing superoxide dismutas	782	782	76%	0.0	95%	
AL939113.1	Streptomyces coelicolor A3(2) complete genome; segment 10/29	782	782	93%	0.0	90%	
AF099014.1	Streptomyces coelicolor strain A3(2) transposase (tnpA) and Fe-containing	782	782	93%	0.0	90%	G
AF012087.1	Streptomyces coelicolor superoxide dismutase (sodF) gene, complete cds	782	782	93%	0.0	90%	
CP002475.1	Streptomyces flavogriseus ATCC 33331, complete genome	780	780	96%	0.0	89%	
FN554889.1	Streptomyces scabiei 87.22 complete genome	780	780	97%	0.0	89%	
CP003275.1	Streptomyces hygroscopicus subsp. jinggangensis 5008, complete genome	752	752	91%	0.0	90%	
FR845719.1	Streptomyces venezuelae ATCC 10712 complete genome	737	737	92%	0.0	89%	
CP002993.1	Streptomyces sp. SirexAA-E, complete genome	728	728	94%	0.0	88%	
EU492452.1	Streptomyces peucetius ATCC 27952 superoxide dismutase (sod1) gene, c	682	682	94%	0.0	87%	
JN005773.1	Streptomyces natalensis strain ATCC 27448 SodF (sodF) gene, complete c	671	671	98%	0.0	86%	
BA000030.3	Streptomyces avermitilis MA-4680 DNA, complete genome	665	665	90%	0.0	87%	
CP002994.1	Streptomyces violaceusniger Tu 4113, complete genome	652	652	96%	0.0	86%	
GU383995.1	Streptomyces violaceoruber strain DSM 40049 Fe-containing superoxide di	643	643	76%	0.0	90%	
GU383994.1	Streptomyces coelicolor strain NRRL B-12000 Fe-containing superoxide dis	643	643	76%	0.0	90%	
GU383993.1	Streptomyces sp. BTG 4-738 Fe-containing superoxide dismutase (sodF) g	643	643	76%	0.0	90%	
GU383992.1	Streptomyces sp. BTG 4-723 Fe-containing superoxide dismutase (sodF) g	643	643	76%	0.0	90%	
GU383991.1	Streptomyces sp. BTG 723 I Fe-containing superoxide dismutase (sodF) ge	643	643	76%	0.0	90%	
GU383990.1	Streptomyces coelicolor strain M110 Fe-containing superoxide dismutase (643	643	76%	0.0	90%	
GU383989.1	Streptomyces sp. BTG 717 I Fe-containing superoxide dismutase (sodF) ge	643	643	76%	0.0	90%	

Figure A2.1: Superoxide dismutase from *Streptomyces coelicolor* A3(2), it will allow sequencing since it is widely distributed among *Streptomyces* spp. (see section 2.2)

Sigma factor from *Streptomyces scabiei* 87.22

TCAGCCGTGCCGGTGGGCGTTGAGCCGGGCCCTGGCGGGTCAGGTGGTCGCGC
TCGGCGAGGTTGGGCGCCGTGCGCGCCGCCTCGGCGTACAGGCGTGCCGCCGTCG
CCGGATCGCC GTCGCGCTCATGGAGATACGCCGCCACCGCCGTGTGCCGAGGGGT
CGAGTCCGGTACGAG CGCGAGCGCCGCCAGCCCGGCGCGGGGGCCGTCGGCCTC
GCCACGGCCACCGCGCGGTTGAGCCGCACGACCGGACTGTCCGGTCAGCCGCGCG
AGTTCGTTCGTACCACTCGACGATCTGCACCCAGTCGGTCTCCTCGGCCGTGGGAGC
GTCCGCGTGCAGCGCGGCGATGGCGGCCTGGGCCTGGAAGTCCGCCAGCCGGTCA
CGGGCGAGCGCCTGCTGCAGGATGCCGATGCCCTCGGCGATGGACGCGGTGTCCC
ACCGGCCCGGTCTGCTCGGCGAGCGGCACCAGGCTGCCGTCGGGCGCGGTACG
GGCTGCCCGCCGGGCATGGTGCAGCAGCATGAGGGCGAGCAGTCC CGCCACCTC
GGGATGGTTCGATCGCGGCCGCCAGCTGCCGGGTGAGCCGGACGGCCTCGGCGGC
GAGGTTCGACGTCCCCGAGTACCCCTCGTTGAAGACCAGGTACAGGACGCGCAGC
ACGGTGGCGACGTCGCCCGGCCGGTTCGAGGGCGACGCCGGAGACCGTGCGCTTGG
CCCGGCTGATGCGCTGCGCCATGGTTCGCTCGGGCACAGATACGCCCGCGCGAT

CTGGCGGGTGGTCAGCCCGCCGACGGCGCGCAGCGTGAGCGCGACCGCCGACGA
 CGGGGTGAGCGAGGGATGGGCGCACAGGAAGTAGAGCCGGAGCGTGTCGTCCGT
 GCCGGGCGCGGGCCCCGGGCGACGGCTCCTCCTCGACGAGGTCCTCGCGCCGGGCGC
 CGGGCGGCGTCCGCGCGGGTTCGCGTCGAGGAACCGGCGCCAGGCCACGGTGACC
 AGCCAGCCCTTCGCGTCCCGCGGCGGGTTCGCCCCGGCCAGACGCGGAGCGCCTCCA
 GCAGCGCGTCTGCACCGCGTCTCGGCCGCGCGAAGTCGGCTCCGCGGGCGGAC
 GAGGACGGCGAGCACACCGGGGGTGTAGGCCGCGCAGCAGGGCCTCGTCCATCGA
 CGGGGTAC

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
FN554889.1	Streptomyces scabiei 87.22 complete genome	2139	2139	100%	0.0	100%
AM420293.1	Saccharopolyspora erythraea NRRL2338 complete genome	1452	1452	98%	0.0	90%
CP000249.1	Frankia sp. CcI3, complete genome	1424	1424	99%	0.0	89%
BA000030.3	Streptomyces avermitilis MA-4680 DNA, complete genome	1415	1415	98%	0.0	89%
CT573213.2	Frankia alni str. ACN14A chromosome, complete sequence	1389	1389	98%	0.0	89%
CP000820.1	Frankia sp. EAN1pec, complete genome	1380	1380	99%	0.0	88%
CP002047.1	Streptomyces bingchengensis BCW-1, complete genome	1352	1352	98%	0.0	88%
AM238663.1	Streptomyces ambofaciens ATCC 23877 left chromosomal arm	1338	1338	99%	0.0	88%
CP002993.1	Streptomyces sp. SirexAA-E, complete genome	1297	1297	98%	0.0	87%
CP002994.1	Streptomyces violaceusniger Tu 4113, complete genome	1297	1297	99%	0.0	87%
CP003275.1	Streptomyces hygroscopicus subsp. jinggangensis 5008, complete genome	1284	1284	98%	0.0	87%
CP003229.1	Streptomyces cattleya DSM 46488 plasmid pSCATT, complete genome	1280	1280	99%	0.0	87%
F0859184.1	Streptomyces cattleya NRRL 8057 plasmid pSCAT, complete genome	1280	1280	99%	0.0	87%
AL939120.1	Streptomyces coelicolor A3(2) complete genome; segment 17/29	1280	1280	98%	0.0	87%
CP002638.1	Verrucosipora maris AB-18-032, complete genome	1254	1254	97%	0.0	87%
FR845719.1	Streptomyces venezuelae ATCC 10712 complete genome	1230	1337	98%	0.0	86%
FO203431.1	Modestobacter marinus str. BC501 chromosome, complete genome	1221	1221	98%	0.0	86%
AP009493.1	Streptomyces griseus subsp. griseus NBRC 13350 DNA, complete genome	1205	1205	98%	0.0	86%
CP002162.1	Micromonospora aurantiaca ATCC 27029, complete genome	1190	1190	98%	0.0	86%
CP002399.1	Micromonospora sp. L5, complete genome	1173	1173	98%	0.0	86%
CP002475.1	Streptomyces flavogriseus ATCC 33331, complete genome	1162	1162	98%	0.0	85%
CP002896.1	Amycolatopsis mediterranei S699, complete genome	1092	1092	97%	0.0	84%
CP002000.1	Amycolatopsis mediterranei U32, complete genome	1092	1092	97%	0.0	84%
GU564398.1	Nocardia sp. ATCC 202099 nocardiacin I biosynthetic gene cluster, comple	1068	1068	97%	0.0	84%
AP011115.1	Rhodococcus opacus B4 DNA, complete genome	968	968	96%	0.0	83%
AP010968.1	Kitasatospora setae KM-6054 DNA, complete genome	893	893	97%	0.0	81%
CP000667.1	Salinispora tropica CNB-440, complete genome	830	830	94%	0.0	81%
CP000431.1	Rhodococcus jostii RHA1, complete genome	806	806	97%	0.0	80%
CP000850.1	Salinispora arenicola CNS-205, complete genome	728	728	97%	0.0	79%
AP006618.1	Nocardia farcinica IFM 10152 DNA, complete genome	725	725	96%	0.0	79%

Figure A2.2: Sigma factor from *Streptomyces scabiei* 87.22, it will allow sequencing since it is widely distributed among *Streptomyces* spp.

[2Fe-2S]-protein GBAD from *Streptomyces scabiei* 87.22

TCACCCCTGGGTGATGCCGAGGCGCTCGTTCACCCAGTCGTGGAAGGCGCCGATG
TGGTGCTCGCTGGGCACCAGCACACCGCCCTTGGCGTACAGGCGTGAGCTCATGC
CGGGCTGGGTCCGCTCGCAGGCCTCGAAGTCCTGGAGGTTGACCCGGTCTGAAGAG
TTCCACGGACCGGCTGACGTCCTTGCCGCTCTCCACGACGTGCGGCAGGTAGAGC
CAGTCGCACTCCACGATCGTGCGGTCCGCCGCGACCGGGTACATGCGGTGGAAGA
TCACGTGGTCGGGGACGAGGTTGATGAACACCTGCGGCCGGACGGTGATCGCGTA
GTAACGGCGGTCTGGTCTCCGCGACGCC CGGGATACGGTCCAGCCCCTCGGAG
CCGTCGACCGTGAAGCCCTGCACGTCCTCGCCGAA CTCCGCACCGTGGCCCACGT
AGTACTGGGCGGGGTACCCGTCCGCGAACTCCGGCAGCACCTCCGTCAGTTCCGG
ATGGATCGTCGCGCAGTGATAGCACTCCATGAAGTTCTCGATGATCAGCTTCCAGT
TCGCCTTACGTCGTACGTGATCCGACGACCCACCGACAGACCGTCGATGTCGTA
CCGCTCGATCGACTCCACGTCACCGAGGCGGGTGACGACCTCGCCGATGACGTCC
TCCTCGAAGGAGGGCGGGTTCTCGGCGAGGCAGACCCAGACATAACCGAGCCATT
CCCGTACGGCGACACTCACCAGGCCGTA CTCCGGTCCGGCCCACGTGCGGGCATCTT
GGTGAGGTTGGGCGCGGGCACCAGCTTGCCGTTCAAGTTCGTACGTCCAGGCGTGG
TACGGGCACTGGAAGGCCCTCTTGACCTCGCCGGACTCCTCGGTGCACAGCTTGG
CCCCGCGGTGCCGGCAGACGTTGAAGTACGCGCGGATCGAGTTGTCCCGGGCCCCG
GGTCACGAGGATGCTCTCGCGGCCACGTGACGGTGCGGAAGGCACCGGGCTTC
GCCAGCTCGGAGGCCCGGGCGACGCAGAACCACATGGTCTCGAATATGTGCTCCT
GCTCCTGGGCGAAGAACGCCGGATCGGTGTAGGAGGAGCCGGGAAGAGTGGCGA
TCAGGCTGTCGGGCAGGCTGGTTCGAGGTCAC

Accession	Description	Query coverage	Max iden
NC_013929.1	Streptomyces scabiei 87.22 chromosome, complete g	100%	100%
ABJJ02000126.1	Streptomyces sviceps ATCC 29083 cont1.126, whole	97%	87%
AHBF01000066.1	Streptomyces acidiscabies 84-104 contig_66, whole	97%	86%
AGDE01000154.1	Streptomyces chartreusis NRRL 12338 contig25.2, wl	98%	85%
ACEZ01000207.1	Streptomyces viridochromogenes DSM 40736 cont1.2	96%	85%
ACEZ01000208.1	Streptomyces viridochromogenes DSM 40736 cont1.2	98%	83%
NC_003155.4	Streptomyces avermitilis MA-4680, complete genome	92%	79%
NC_009142.1	Saccharopolyspora erythraea NRRL 2338 chromosom	97%	78%
ACEX01000119.1	Streptomyces himastatinicus ATCC 53653 cont1.119,	95%	78%
AGNH01000080.1	Streptomyces sp. TOR3209 Contig80, whole genome	93%	78%
AGDD01004396.1	Streptomyces chartreusis NRRL 3882 contig2548.1, v	49%	87%

Figure A2.3: [2Fe-2S]-protein GBAD from *Streptomyces scabiei* 87.22, it will allow sequencing since it is widely distributed among *Streptomyces* spp.

Pyruvate phosphate dikinase:

TTACCGGTGGTCGCTGCCCTTCGACCGCGACGCCGCCCGGCCCGCCTCCAGGCGG
GCCACCGGGATCCGGAACGGGGAGCAGGAGACGTAGTCCAGACCCACCTCGTGG
AAGAAGTGGACCGACTCCGGGTCGCCGCCGTGCTCGCCGCAGACGCCGAGCTTCA
GGTCGGGACGCGTGGCGCGGCCGGCCTTCGCGGGCGGCGGCGACCAGGGAGCCGA
CACCGTCCTTGTCGATGGTCTCGAACGGGGAGACGCCGAAGATGCCCTTCTCCAG
GTAGGCCGTGAAGAAGCTGGCCTCGACGTCGTCGCGGCTGAAGCCCCACACGGTC
TGGGTCAGGTCGTTGGTGCCGAAGGAGAAGAAGTCCGCGGCCTCGGCGATCTGCC
CGGCGGTCAGCGCGGCGCGCGGCAGCTCGATCATCGTGCCGATCGACAGCTTGAG
CTTGGCGCCGGTGGCGGCCTCGACCTCGGCGATGACCTGGTCCGGCCTCCTCGCGG
ACGATCTCCAGCTCCTGGACGGTGCCGACCAGCGGGATCATGATCTCGGCGCGCG
GGTCGCCCTTGGCCGCCTTGCCTCGGCCCGGCCTCGGCGATGGCGCGGACCTG
CATGGTGAACAGGCCGGGGATGACCAGGCCGAGGCGCACCCCGCGCAGGCCAG
CATCGGGTTCTGCTCGTGCAGCCGGTGCACGGCCTGGAGCAGGCGCAGCTCGTTC
TCGTGCGGCTCCTGACGGGACTCCGCGAGCGCCACACGTACCGACAGTTCGGTGA
TGTCGGGCAGGAACTCGTGCAGCGGCGGGTTCGAGGAGGCGGACCGTGACCGGGA
GGCCGTCCATCGACTCGAAGAGCTGGACGAAGTCCTGCTTCTGGAGCGGCAGCAG
CTCCTTGAGGGACTCCTCGCGCACCGCCTCGGTGTCGGCGAGGATCAGCCGCTCC
ACCAGCTCGCGCCGGTCCCGGAGGAACATGTGCTCGGTGCGGCACAGCCCGATGC

CCTGGGCACCGAACCGGGCGGGCGCGCAGCGCGTCCTCGGGCGTTGTCGGGCGTTGGC
GCGCACCCGCAGCCGGCGCTTGCGGTTCGGCGAAGGCCATGATCCGGTGCACGGCC
TCGACCAGCTCGTCGGCGTCCTCGGGCGCCCGGTGCATCCGGCCCTCGAAGTACTC
GACGACCGGGGAGGGCACGACCGGCACCTCGCCGAGGTAGACCTTGCCGGAGGA
GCCGTTCGATGGAGATCACGTTCGCCCTCCTCGACGACGTGCCCGCCGGGCACGGTC
ATCCGGCGGGCGCTTGGTGTTCGACCTCCAGTTCCCTCGGGCGCCGCAGACACAGGTCTT
GCCCATGCCGCGCGCCACCACGGCCCGGTGGGAGGTCTTGCCGCCGCGCGAGGTC
AGGATGCCCTCGGGCGGCGATCATGCCGTCCAGGTCGTCGGGGTTGGTCTCGCGGC
GGACCAGGATGACCTTCTCGCCGGACCGGGACCACTTGACGGCGGTGTACGAGTC
GAAGACGGCCTTGCCGACCGCCGCGCCCGGCGAGGCGGCGATGCCCCGGCCGAC
CTTCTCGACCTTCGCCTCGTCGTCGAAGCGGGGAACATCAGCTGGGCGAGCTGG
GCGCCGTTGACCCGCTGGAGGGCCTCGGCCTCGTCGATCAGGCCCTGGTTCGACGA
GCTGGGTGGCGATCCGGAAGGCGGCGCCCGCGGTGCGCTTGCCGACGCGGGTCTG
GAGCATCCACAGCTGACCGCGCTCGATGGTGAACCTCGATGTTCGCACAGGTCCAGG
TAGTGGTTCTCCAGCGTCTCCATGATCTGCATCAGCTGGTTCGTACGACTTCTTGTC
GATCTGCTCCAGGTCCGCCAGCGCGACGGTGTGCGGATGCCCGCGACGACGTCC
TCGCCCTGGGCGTTCTGGAGGTAGTCGCCGTAGACGCCCTGGTGCCCGGAGGCGG
GGTCGCGGGTGAAGGCGACGCCGGTGCCGGAGTCGGGGCCGAGGTTGCCGAAGA
CCATGGAGCACACGTTGACGGCGGTGCCGAGGTCGTGCGGGATGCGCTCCTGGCG
GCGGTAGAGCTTGGCGCGCTCGCCGTTCCAGGACTCGAAGACCGCCTTGATGGCG
AGGTCCATCTGCTCGCGCGGGTCCTGCGGGAAGTCCCGGCCGGCCTCGGTCTTGA
CGATCTTCTTGAACCTGGTGACGAGCTTCTTGAGGTCGGCGGCCTCCAGCTCGGTG
TCGACCGCGACCTTCTTGGCCGTCTTGGCCGCCTCCAGCGCCTCCTCGAAGAGGTC
GCCGTTCGACGCCGAGGACGGTCTTGCCGAACATCTGGATCAGGCGGCGGTAGGAG
TCCACGCGAAGCGGTTCGTCGCCGGCCTGCTTGGCCAGGCCCTGCACGGACTTGT
CGGAGAGGCCGATGTTGAGGACGGTGTCCATCATGCCGGGCATGGAGAACTTGGC
CCCGGAGCGGACCGACACGAGGAGGGGGTCGTCGGCCTGGCCGAGCTTCTTGCCC
ATCGTGGCTTCCAGGGCGTCGAGGTGCGCACTCACCTCGTCACGCAGTCCCGCGG
GCTCCTCGCCGCTGTCGAGGTAGGTCTTGCAGGCCTCGGTGGTGATCGTGAAGCC
GGGAGGGACCGGGAGACCCAGGTTGGTCATCTCGGGCGAGGTTGGCGCCCTTGCCG

CCAGGAGGTCCTTGAGGTCCTTGTTGCCCTCGGTGAAGTCGTAACGAACCTTCGCT
GCGTTCGCTACCTGGGGATCTTTGTTTTCCGACAC

Sequences producing significant alignments:						
Accession	Description	Max score	Total score	Query coverage	E value	Max ident
NC_003888.3	Streptomyces coelicolor A3(2) chromosome, complete genome	5035	6236	100%	0.0	99%
ACEY01000224.1	Streptomyces lividans TK24 cont1.224, whole genome shotgun sequence	5001	5001	100%	0.0	99%
AHGS01000047.1	Streptomyces coelicoflavus ZG0656 contig2083, whole genome shotgun seq	4739	4739	99%	0.0	98%
ASHH01000187.1	Streptomyces sp. TOR3209 Contig187, whole genome shotgun sequence	4362	4362	99%	0.0	96%
ACUR01000173.1	Streptomyces sp. ei4 cont1.173, whole genome shotgun sequence	4082	4142	99%	0.0	97%
AGDE01000134.1	Streptomyces chartreusis NRRL 12338 contig19.9, whole genome shotgun s	3916	3916	98%	0.0	93%
ACEZ01000105.1	Streptomyces viridochromogenes DSM 40736 cont1.105, whole genome sho	3916	3916	98%	0.0	93%
ACFA01000617.1	Streptomyces griseoflavus Tu4000 cont1.617, whole genome shotgun sequ	3890	3890	98%	0.0	93%
ABYA01000392.1	Streptomyces ghanaensis ATCC 14672 cont1.392, whole genome shotgun s	3860	3860	98%	0.0	93%
AGBF01000189.1	Streptomyces zinciresiatens K42 contig00183, whole genome shotgun sequ	3790	3790	98%	0.0	92%
ABJ02000245.1	Streptomyces avicues ATCC 29083 cont1.245, whole genome shotgun sequ	3690	3690	98%	0.0	91%
AEYX01000002.1	Streptomyces griseourantiacus M045 Contig002, whole genome shotgun se	3677	3677	98%	0.0	91%
NC_017765.1	Streptomyces hygroscopicus subsp. jinggangensis 5008 chromosome, comp	3620	5155	98%	0.0	91%
NC_013929.1	Streptomyces scabiei 87.22 chromosome, complete genome	3555	4703	98%	0.0	91%
AJJM01000004.1	Streptomyces somaliensis DSM 40738 contig004, whole genome shotgun se	3546	3546	98%	0.0	91%
NC_003195.4	Streptomyces avermitilis MA-4680, complete genome	3513	3568	99%	0.0	100%
ABYC01000167.1	Streptomyces albus J1074 cont3.167, whole genome shotgun sequence	3500	3500	98%	0.0	90%
AHBF01000051.1	Streptomyces acidificabies 84-104 contig_51, whole genome shotgun seque	3496	3496	98%	0.0	90%
AEVW01002230.1	Streptomyces sp. PP-C42 contig2230, whole genome shotgun sequence	3445	3445	98%	0.0	90%
ABJI02000543.1	Streptomyces pristinaespiralis ATCC 25486 cont1.543, whole genome shotg	3413	3413	98%	0.0	90%
ABJF01000111.1	Streptomyces sp. Mgi cont1.111, whole genome shotgun sequence	3367	3367	98%	0.0	89%
ABYX01000096.1	Streptomyces roseosporus NRRL 11379 cont3.96, whole genome shotgun se	3339	3339	98%	0.0	89%
ABYB01000098.1	Streptomyces roseosporus NRRL 15998 cont3.98, whole genome shotgun se	3339	3339	98%	0.0	89%
AGSW01000082.1	Streptomyces sp. W007 contig00085, whole genome shotgun sequence	3328	3328	98%	0.0	89%
ACEW01000238.1	Streptomyces sp. C strain C cont1.258, whole genome shotgun sequence	3328	3328	98%	0.0	89%
AJUD01000192.1	Streptomyces globisporus C-1027 Scaffold32_2, whole genome shotgun seq	3290	3290	98%	0.0	89%
NC_010572.1	Streptomyces sp. subsp. griseus NBRC 13350 chromosome, complete g	3254	3254	97%	0.0	89%
ADCF02000002.1	Streptomyces griseus XylebKG-1 scaffold_1_Cont2, whole genome shotgun i	3238	3238	97%	0.0	89%
AGDF01000417.1	Streptomyces sp. Wigar10 contig00423, whole genome shotgun sequence	3216	3216	97%	0.0	89%
NC_015953.1	Streptomyces sp. SirexAA-E chromosome, complete genome	3190	3190	98%	0.0	88%
AJSZ01000677.1	Streptomyces tsukubaensis NRRL18488 Contig677, whole genome shotgun	3171	3171	99%	0.0	88%
ACEX01000540.1	Streptomyces himastatinicus ATCC 53653 cont1.540, whole genome shotgun	3136	3136	97%	0.0	88%
AFRF01001818.1	Streptomyces xinghaiensis S187 contig_1924_1, whole genome shotgun seq	3099	3099	97%	0.0	88%
AGV010001427.1	Streptomyces auratus AGR0001 Scaffold1_127, whole genome shotgun sequ	3094	3094	98%	0.0	88%
ABJH01000120.1	Streptomyces clavuligerus ATCC 27064 cont1.120, whole genome shotgun s	3085	3085	97%	0.0	88%
NC_016582.1	Streptomyces bingchengensis BCW-1 chromosome, complete genome	3081	3081	97%	0.0	88%
NC_015957.1	Streptomyces violaceusniger Tu 4113 chromosome, complete genome	3081	3081	97%	0.0	88%
ACEU01000159.1	Streptomyces sp. SPB78 cont1.159, whole genome shotgun sequence	3027	3027	97%	0.0	87%
ADXB01000447.1	Streptomyces sp. SA3_actf contig00139, whole genome shotgun sequence	3011	3011	97%	0.0	87%
ADXA01000032.1	Streptomyces sp. SA3_actG contig00081, whole genome shotgun sequence	3011	3011	97%	0.0	87%
NC_016114.1	Streptomyces flavogriseus ATCC 33331 chromosome, complete genome	3001	3001	98%	0.0	87%
AFH01000002.1	Streptomyces sp. Tu6071 contig2, whole genome shotgun sequence	2977	2977	97%	0.0	87%
ABIG02000604.1	Streptomyces sp. SPB74 cont3.604, whole genome shotgun sequence	2953	2953	97%	0.0	87%
NC_017586.1	Streptomyces cattleya NRRL 8057 = DSM 46488 chromosome, complete ger	2944	2944	97%	0.0	87%

Figure A2.4: Pyruvate phosphate dikinase from *Streptomyces coelicolor* A3(2), it will allow sequencing since it is widely distributed among *Streptomyces* spp.

Amino Acid Adenylation Domain Protein

TCACACCCCTTCGGTCAGGGCACGGATCGTCGGTGTGCGGTAGAGGTCGCGCAAG
GGCAGACGGCAGGTCGCGGGCGCGAAGCGCGGCGCTGACCGGAACGGCGAAGAG
CGAGTTGCCGCCAGTTTCGAAGAAGTCGTCGTCCAGGCCACGGGCGTGCCGAGC
ACGGTCGCCCAGATCTCCAGAAGAGCCTCGGGCAGGCCGGCGGGCCGCCCGGAG
CCCGCGGCCGGCTGCCGGGGGGCGGCGGAGGACCGGGACGGGGCGGGGAGACGG
GCGGTGTCGAGTTTGCCGTTTCGTCGTGAGGGGGAGCGCGTCCACGGCGGTGACCG
TGAGGGGACCATGTGGCCGGGCAGCAGCCCCGCGCTGCGTTTGCGGATGCCGGG
GAGGTCGGCCAGGCCCTCCGCGGGTACGACGTAGGCGTCCAGCCGGGCGGGCGGC
CGGGTCGTCGGGATCGTCGCGGCACGACCACCGCGGCCGCGCGGACACCGGG
GTCTCCAGCAGTACGGAGCGGATCTCGTCGAGCTCGATGCGGAAGCCGCGGATC

TTGACCTGGCTGTCGATCCGTCCGAGGTGTTCCAGCCTGCCGTCCGGCAGCAGCCG
CCCTAGGTCCCCGCTCCGGTACATCCGTCCGCCGGTGACGGGATCGGTACAGAAA
CGCTGCGCGGTTCAGGTCCGGCCGGCCGAGGTAGCCGGTGGCGACGCCGGCCCCGC
CCACGCAGATCTCGCCGGTACCCCCGGCGGTACCGGTCTGCCGTGCGTGTCCGT
GACGTACAGGTGCCAGCCGGGGATGGCCGGCCCCACGGAGCGTGAGCCGGCCAG
GGCGAGTTCGCGGGTGACGGTCTGGGCGGTGACGTGCACGGTTCGTCTCCGTGATG
CCGAACATGTTGACCATGCGGCAGGCGTCTCCGGATGCCGGTTCGAACCAGGGCA
GCAGCATCCGGCTGTCGAGCGGCTCGCCTCCGAACACCACGAGCCGGACAGCGAG
ACCGGCGTGGTTCGACCCGGAGCAACTGCGAGAAGGCGGACGGTGTCTGGCTCAG
GACGGTGACCTGTTCCGGCGACCAGCAGGTTCGCGGAACTCCTCGGGGTCGCGCGAG
ACGAAGTAGGGCACGACGACGAGCCGTCCGCCCGTCAGCAGGCAGCCCCAGATC
TCCCACACGGAGAAGTCGAAGGCGCCGGAGTGGAACAGCGTCCACACGTCGCCG
GGGCCGAGCCGGTACGTGTCACGGGTGGCCTCGATGAGCGAGACGACGTTGCCGT
GCGGGACGACGACACCCTTGGGGCGCCCGGTGGAGCCCGAGGTGTAGATGACGT
ACGCGGGATCGTCCGGGCCCCGGCGCCACCTCCGGTGCCGACGGCCGGTCCGGGCGC
CGCCTCCAGGAGTTCACCGGGTGTACGCGCGGTGGTGCCGTCCCGCGCGGGGAAC
TCGGGCAGCCGCGTGACCACCACGCGCAGGCCGGCGTCCGACGCCGTCCAGGCC
AGGCGGTCCGGCGGGATAGGCCGGGTCCGGTCCGTACGTACGCGGCTCCCGCCTTGA
GCACGCCGAGCATGGTGACGACCAGTTCGACGGTCCGTTCCAGGCAGACACCGAC
CCGGTCCCGGTGCCGGACCCCGTGCGCCCGGAGGCCGGCGGCGAGGAGAGTGCC
CCTGTCGTCCAGCTCCCGGTAGGTGAGGCTCACGTCACCGTCCGCTACCGCGACG
GCGTCCGGGGTGGCGGACACGGCTCGCCGGAACGCGTCCAGCAGCGAGATCCCC
GCGGCGGCGACGTCCCCGTGGGGGCGGCCGAGCCGTGCGATCCGGTCCCGCTCGG
CGGGGTCCAGCAGGTCCGCGTCCCGACGCGGGACTGCGGCGCGCGCGCCGACA
GCGCGTACACCCGGGCCAGGTGCCCGGCGAACTGCTCGACGGCCTCGGGGGAGA
AGTGGCTGCGGAGGTAGGAGAACCACAGGCGTGAACCCCGTCCGGCAGAGGGA
CGAGCACGGTCGTCAGCGGGAACGCCGGAGCGAGGAGCGGCAGGTAGTCGTCAC
CGGTCCGGCCCGTCCGGGCGGATCCGTGAGCAGCCCGGCGGTGACGGGCCGGTC
CACCCGGACGGGGGCGGACTCGCACCGAGCCCTCAGGTCCGCGAGGGTCCGCCG
CGGATCCACCGTGATCTCCCGCGCGATGCCACCGGGACCGGCGACGGTACCCCC
TCCGTCCGCGCCGTGCCCGTTCGTAGCGGCTCAGGGTCAGGGCGAGAGCGGCCGTCA

GCAGCCCGGGGCCGCCCGCCGGCTCCTGGGGCAGGTCCAGGTGGTACGTGCCGTG
GCGGGACGAACCGGGCTCGCCCCGGCCCCAGTCCGGCGCCGGACCCTCGGCCGGC
CGGGCGACACGATCCCCCTCGGGTGCGGGACACGGCCCCGGCGGCCCGGGTCCCCG
CCAGCATCGCGGTCAGTGCGCCCAGACCGCCCCGGTCCAGCCGGTCCCGGTCCGC
CACGACCACCAGGTTCGGCCTCCCCGTTCGGCGTACCGCAGGAGGACACACCGCACG
CCGGCCCGCCGGCCGCGCATCTCGGCGTCGAGACGGCGCTCGGCGAGCGGCTCGG
CCGCGCCCCGCCGGCACGTCGTCCACCCACAGCCGCCAGGGCCCCGGCGGGGACA
GCCGTTCGGCGAGTGTCTGTCCGGCACCGGAGGGCCGGTCAGCCGGACGCGGAC
CGCATGGCACTGTGCACCCAGCAC

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
NC_016114.1	Streptomyces flavogriseus ATCC 33331 chromosome, complete genome	4479	4896	96%	0.0	100%
NC_013929.1	Streptomyces scabiei 87.22 chromosome, complete genome	850	1112	54%	0.0	96%
AJJMO1000044.1	Streptomyces somaliensis DSM 40738 contig044, whole genome shotgun sequence	444	444	38%	1e-120	75%
ABYA01000114.1	Streptomyces ghanaensis ATCC 14672 cont1.114, whole genome shotgun sequence	239	380	29%	5e-59	73%
ACFA01000901.1	Streptomyces griseoflavus Tu4000 cont1.901, whole genome shotgun sequence	217	217	29%	2e-52	73%
NC_003888.3	Streptomyces coelicolor A3(2) chromosome, complete genome	211	286	30%	1e-50	87%
NC_015957.1	Streptomyces violaceusniger Tu 4113 chromosome, complete genome	207	266	30%	1e-49	100%
ACFA01000644.1	Streptomyces griseoflavus Tu4000 cont1.644, whole genome shotgun sequence	200	200	31%	2e-47	72%
AGSW01000170.1	Streptomyces sp. W007 contig00175, whole genome shotgun sequence	180	180	17%	3e-41	74%
ABYX01000056.1	Streptomyces roseosporus NRRL 11379 cont3.56, whole genome shotgun sequence	178	178	11%	1e-40	77%
ABYB01000050.1	Streptomyces roseosporus NRRL 15998 cont3.50, whole genome shotgun sequence	178	178	11%	1e-40	77%
ADGX01000035.1	Frankia sp. EUN1f ctg00311, whole genome shotgun sequence	145	145	30%	1e-30	71%
NC_013440.1	Haliangium ochraceum DSM 14365 chromosome, complete genome	145	503	29%	1e-30	87%
ACEY01000183.1	Streptomyces lividans TK24 cont1.183, whole genome shotgun sequence	143	143	25%	4e-30	72%
AFWL01000308.1	Ralstonia solanacearum Y45 plasmid megaplasmid plasmidContig3, whole genome	128	128	9%	1e-25	76%
NC_013235.1	Nakamurella multipartita DSM 44233 chromosome, complete genome	128	251	11%	1e-25	75%
AAKL01000003.1	Ralstonia solanacearum UW551 Cont0580, whole genome shotgun sequence	124	384	10%	1e-24	79%
NC_003296.1	Ralstonia solanacearum GM1000 plasmid pGM1000MP, complete sequence	122	311	9%	5e-24	76%
NC_008095.1	Myxococcus xanthus DK 1622 chromosome, complete genome	121	186	8%	2e-23	93%
NC_013510.1	Thermomonospora curvata DSM 43183 chromosome, complete genome	115	115	8%	9e-22	77%
ACEX01000635.1	Streptomyces himastatinicus ATCC 53653 cont1.635, whole genome shotgun sequel	110	201	9%	4e-20	75%
NC_017575.1	Ralstonia solanacearum Po82 megaplasmid, complete sequence	102	194	7%	7e-18	76%
NC_016582.1	Streptomyces bingchengensis BCW-1 chromosome, complete genome	102	370	4%	7e-18	93%
AGDF01000005.1	Streptomyces sp. Wigar10 contig00005, whole genome shotgun sequence	100	100	8%	3e-17	75%
AGBF01000316.1	Streptomyces zinciresistens K42 contig00316, whole genome shotgun sequence	100	100	11%	3e-17	73%
ADNW01000063.1	Mycobacterium parascrofulaceum ATCC BAA-614 contig00076, whole genome shotgu	100	100	6%	3e-17	77%
NC_018750.1	Streptomyces venezuelae ATCC 10712, complete genome	100	190	11%	3e-17	96%
NC_006361.1	Nocardia farcinica IFM 10152 chromosome, complete genome	100	393	15%	3e-17	95%
AHGS01001576.1	Streptomyces coelicoflavus ZG0656 contig01697, whole genome shotgun sequence	99.0	99.0	7%	9e-17	76%
AGDC01000417.1	Streptomyces lysosuperificus ATCC 31396 contig96.1, whole genome shotgun sequel	97.1	97.1	2%	3e-16	90%
AGDF01000056.1	Streptomyces sp. Wigar10 contig00057, whole genome shotgun sequence	95.3	95.3	2%	1e-15	92%
AGBF01000050.1	Streptomyces zinciresistens K42 contig00050, whole genome shotgun sequence	95.3	175	10%	1e-15	93%
AJLQ01000109.1	Rhodococcus sp. DK17 ctg170000109, whole genome shotgun sequence	93.5	166	7%	4e-15	89%
ABYX01000031.1	Streptomyces roseosporus NRRL 11379 cont3.31, whole genome shotgun sequence	93.5	93.5	4%	4e-15	82%
ABYB01000020.1	Streptomyces roseosporus NRRL 15998 cont3.20, whole genome shotgun sequence	93.5	164	10%	4e-15	82%
ABJH01000401.1	Streptomyces clavuligerus ATCC 27064 cont1.401, whole genome shotgun sequence	93.5	241	7%	4e-15	90%
NC_008268.1	Rhodococcus jostii RHA1 chromosome, complete genome	93.5	305	7%	4e-15	89%
NC_013595.1	Streptosporangium roseum DSM 43021 chromosome, complete genome	91.6	91.6	6%	2e-14	77%
AGSW01000095.1	Streptomyces sp. W007 contig00099, whole genome shotgun sequence	89.8	170	2%	5e-14	92%
ACEZ01000045.1	Streptomyces viridochromogenes DSM 40736 cont1.45, whole genome shotgun sequ	89.8	89.8	2%	5e-14	90%
AFIF01000147.1	Mycobacterium avium subsp. paratuberculosis S397 MavS397DRAFT_contig00133c,	87.9	87.9	8%	2e-13	74%
ACBV01000013.1	Mycobacterium kansasii ATCC 12478 contig00092, whole genome shotgun sequence	86.1	86.1	6%	7e-13	76%
ABYC01000412.1	Streptomyces albus J1074 cont3.412, whole genome shotgun sequence	86.1	86.1	3%	7e-13	85%
NC_018581.1	Gordonia sp. KTR9 chromosome, complete genome	86.1	170	6%	7e-13	94%

Figure A2.5: Amino Acid Adenylation Domain Protein from *Streptomyces flavogriseus* ATCC 33331, it will allow sequencing since it is widely distributed among *Streptomyces* spp.

Transcriptional regulatory:

CTACCGCTCCGGCGGTGCGAAGAGGTCGTCCTGGGCGCGGTTCGCGAGCGGTGAGC
AGGGCGCCGCGCAGTACCGCGCCGCCGCCCAGGGTGCTGGGCCGTACCTCCGTGG
CCAGCGGTGACATGCGGGTGAGCCGGTGCTGGACGCGGTTCGGCGAGGGCGTCGG
CGCCGGCCCCGGCCGATCTCACCGGCGAGCACCAGGCAGCCAGGGTTCGAGGACGG
AGACGACGGCGGCGGCGCCGAGGACGACCCGGTTCGGCGAGCGCGTCGAGGAACC
GCTCGGCGGCGGGGTTCGCTCCCGGACGCCGACCGGCGTACCGCCTCCCGGACCAG
CGCGGCGGCGGTGCGGCTCGGGCCCCGGCGGGTTCGTGCCGTCACGCCGTGCTCCGCC
GCGAGGGCGGTGACCGCCGCGCGCCCGCAACGGAGTGGAAGCCGCGCTCGCAG
TCGGTGGCGGAGGGCAGGCCGGCGGTGCCCGGCACCGGCATGAAGCCGATCTCG
CCGGCGCCCCCGGACACGCCCCGGCGCAGGGTGCCGTCGAGGACGACGGCGGCG
CCGATGCCGAGGCCGAGCCAGAGCAGGACGAAGGTGTCCCGGTCCCGGGCGGCG
CCGTTCGCGCTGTTTCGGCCAGGGCGGCGAGGTTGGTCTCGTTCTCGACGCCGACCC
GGGCCTCGGGGAACTTCTCCTGGAGCGCGGCCATCAGCCGACGATGCCACTCCGG
CAGCCCCGGTGGAGTCCCGCAGTTCACCGCCGGCCGGGTTCGATCAGGCCGGGCGCG
CCGATGCCGACGGTGTGCAGCCGGTTCGGCGCCGGCCTCCTTGACCGCGCGCTCGA
CCAGGCTCACCGCCTGCTCCACCGCGGGACCGGTGCCGGAGTGGTCGTCGATGGG
CACGGACGCCTCGGCGAGGACCCGGCCGAGCAGGTTCGGAGACGAGCACGGCGAC
GCCTTCGGTGCACGTCGAGGGCGGCGAGGTAGGCGCGCTCGGCGACGATGCCG
TACAGCTTGGCGTTGGGTCCGCGGGCGCTGCTCGCCCCGACTCCCCCGCCACCTCGAT
GAGTCCGGAGGCGGTGAGGCGTTCGACGAGGTTCGGCGACGGTGGGCCGGGACAG
GCCGGTCAGCTGCTTCAACTGCCCTGCCGTCAACGGGCCCTCCCGCTGGAGCAGA
TGAAGGGCGAGCCGGTCGTTGATGGCCCCGGGCGGTGCTCGGTGATGCGGGCAT

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
NC_003888.3	Streptomyces coelicolor A3(2) chromosome, complete genome	2217	2217	100%	0.0	100%
ACEY01000271.1	Streptomyces lividans TK24 cont1.271, whole genome shotgun sequence	2194	2194	100%	0.0	99%
AHGS01001380.1	Streptomyces coelicoflavus ZG0656 contig01437, whole genome shotgun sequence	1851	1851	100%	0.0	95%
AGNH01000085.1	Streptomyces sp. TOR3209 Contig85, whole genome shotgun sequence	1428	1428	99%	0.0	88%
ABYA01000489.1	Streptomyces ghanaensis ATCC 14672 cont1.489, whole genome shotgun sequence	979	979	98%	0.0	82%
NC_003155.4	Streptomyces avermitilis MA-4680, complete genome	961	961	98%	0.0	82%
AGBF01000005.1	Streptomyces zinciresistens K42 contig00005, whole genome shotgun sequence	880	880	99%	0.0	80%
NC_013929.1	Streptomyces scabiei 87.22 chromosome, complete genome	835	835	99%	0.0	80%
AGDE01000200.1	Streptomyces chartreusis NRRL 12338 contig35.2, whole genome shotgun sequence	832	832	65%	0.0	86%
ACEZ01000055.1	Streptomyces viridochromogenes DSM 40736 cont1.55, whole genome shotgun sequence	797	797	65%	0.0	85%
ABJ02000149.1	Streptomyces sviveus ATCC 29083 cont1.149, whole genome shotgun sequence	774	774	74%	0.0	82%
AGDD01002889.1	Streptomyces chartreusis NRRL 3882 contig1482.1, whole genome shotgun sequence	752	752	61%	0.0	85%
AEYX01000039.1	Streptomyces griseoaurantiacus M045 Contig039, whole genome shotgun sequence	708	708	98%	0.0	78%
NC_018750.1	Streptomyces venezuelae ATCC 10712, complete genome	665	665	98%	0.0	77%
AGDF01000279.1	Streptomyces sp. Wigar10 contig00285, whole genome shotgun sequence	553	553	97%	7e-154	76%
ACUR01000050.1	Streptomyces sp. e14 cont1.50, whole genome shotgun sequence	544	544	48%	4e-151	84%
ABYC01000074.1	Streptomyces albus J1074 cont3.74, whole genome shotgun sequence	521	521	88%	2e-144	76%
NC_016114.1	Streptomyces flavogriseus ATCC 33331 chromosome, complete genome	507	507	98%	6e-140	75%
ABYX01000051.1	Streptomyces roseosporus NRRL 11379 cont3.51, whole genome shotgun sequence	499	499	99%	9e-138	75%
ABYB01000036.1	Streptomyces roseosporus NRRL 15998 cont3.36, whole genome shotgun sequence	499	499	99%	9e-138	75%
ADFC02000002.1	Streptomyces griseus XylebKG-1 scaffold_1_Cont2, whole genome shotgun sequence	475	475	96%	2e-130	75%
NC_010972.1	Streptomyces griseus subsp. griseus NBRC 13350 chromosome, complete genome	464	464	96%	3e-127	74%
ADXAO1000008.1	Streptomyces sp. SA3_actG contig00122, whole genome shotgun sequence	459	459	98%	2e-125	74%
AGDC01000808.1	Streptomyces lysosuperificus ATCC 31396 contig200.4, whole genome shotgun sequence	416	416	69%	1e-112	76%
NC_017765.1	Streptomyces hygrosopicus subsp. jinggangensis 5008 chromosome, complete genome	412	412	34%	1e-111	85%
ADXBO1000464.1	Streptomyces sp. SA3_actF contig00892, whole genome shotgun sequence	375	375	73%	2e-100	75%
AEWS01003913.1	Streptomyces sp. PP-C42 contig3913, whole genome shotgun sequence	178	178	19%	5e-41	81%
NC_016109.1	Kitasatospora setae KM-6054, complete genome	135	135	25%	3e-28	75%

Figure A2.6: Transcriptional regulatory from *Streptomyces coelicolor* A3(2), it will allow sequencing since it is widely distributed among *Streptomyces* spp.

rDNA (16 S) from *Streptomyces coelicolor* A3(2)

CATTCACGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACA
CATGCAAGTCGAACGATGAACCACTTCGGTGGGGATTAGTGGCGAACGGGTGAGT
AACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAA
TACCGGATACTGACCCTCGCAGGCATCTGCGAGGGTCGAAAGCTCCGGCGGTGAA
GGATGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGAC
GACGGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCC
CAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTG
ATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGC
AGGGAAGAAGCGAAAGTGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTG
CCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTGTCCGGAATTATTGGGCGTA
AAGAGCTCGTAGGCGGCTTGTCACGTTCGGTTGTGAAAGCCCGGGGCTTAACCCCG
GGTCTGCAGTCGATACGGGCAGGCTAGAGTTCGGTAGGGGAGATCGGAATTCTG
GTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGAT
CTCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTA
GATACCCTGGTAGTCCACGCCGTAACGGTGGGCACTAGGTGTGGGCAACATTCC
ACGTTGTCCGTGCCGACGCTAACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCG
CAAGGCTAAAACCTCAAAGGAATTGACGGGGGCCCCGCACAAGCGGCGGAGCATGT
GGCTTAATTCGACGCAACGCGAAGAACCTTACCAAGGCTTGACATACACCGGAAA
GCATCAGAGATGGTGCCCCCTTGTGGTTCGGTGTACAGGTGGTGCATGGCTGTCC
TCAGCTCGTGTCGTGAGATGTTGGGTAAAGTCCCAGCAACGAGCGCAACCCTTGTCC
CGTGTTGCCAGCAAGCCCTTCGGGGTGTGGGGACTCACGGGAGACCGCCGGGGT
CAACTCGGAGGAAGGTGGGGACGACGTCAAGTCATCATGCCCTTATGTCTTGGG
CTGCACACGTGCTACAATGGCCGGTACAATGAGCTGCGATAACCGCAAGGTGGAGC
GAATCTCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGA
AGTCGGAGTCGCTAGTAATCGCAGATCAGCATTGCTGCGGTGAATACGTTCCCGG
GCCTTGTACACACCGCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCGGT
GCCCAACCCCTTGTGGGAGGGAGCTGTCGAAGGTGGGACTGGCGATTGGGACGA
AGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCT (54).

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
NC_003888.3	Streptomyces coelicolor A3(2) chromosome, complete genome	2828	16931	100%	0.0	100%
AHGS01000024.1	Streptomyces coelicoflavus ZG0656 contig2022, whole genome shotgun sequence	2780	2780	100%	0.0	99%
AHGS01000042.1	Streptomyces coelicoflavus ZG0656 contig2072, whole genome shotgun sequence	2780	2780	100%	0.0	99%
AHGS01000056.1	Streptomyces coelicoflavus ZG0656 contig2096, whole genome shotgun sequence	2780	2780	100%	0.0	99%
AHGS01000064.1	Streptomyces coelicoflavus ZG0656 contig2114, whole genome shotgun sequence	2780	2780	100%	0.0	99%
AHGS01001749.1	Streptomyces coelicoflavus ZG0656 contig01753, whole genome shotgun sequence	2780	2780	100%	0.0	99%
AHGS01000038.1	Streptomyces coelicoflavus ZG0656 contig2067, whole genome shotgun sequence	2774	2774	100%	0.0	99%
AGDD01000151.1	Streptomyces chartreusis NRRL 3882 contig56.2, whole genome shotgun sequence	2699	2699	100%	0.0	98%
ADXA01000222.1	Streptomyces sp. SA3_actG contig00029, whole genome shotgun sequence	2669	2669	100%	0.0	98%
AGBF01000432.1	Streptomyces zinciresistens K42 contig00432, whole genome shotgun sequence	2658	2658	100%	0.0	98%
AFHJ01000001.1	Streptomyces sp. Tu6071 contig1, whole genome shotgun sequence	2658	2658	100%	0.0	98%
AFHJ01000002.1	Streptomyces sp. Tu6071 contig2, whole genome shotgun sequence	2658	13292	100%	0.0	98%
AGDE01000038.1	Streptomyces chartreusis NRRL 12338 contig4.8, whole genome shotgun sequence	2654	2654	100%	0.0	98%
NC_017765.1	Streptomyces hygrosopicus subsp. jinggangensis 5008 chromosome, complete genome	2615	15657	100%	0.0	98%
NC_003155.4	Streptomyces avermitilis MA-4680, complete genome	2610	15662	100%	0.0	97%
NC_016582.1	Streptomyces bingchengensis BCW-1 chromosome, complete genome	2601	15588	100%	0.0	97%
AJSZ01000237.1	Streptomyces tsukubaensis NRRL18488 Contig237, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AJSZ01000417.1	Streptomyces tsukubaensis NRRL18488 Contig417, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AJSZ01000510.1	Streptomyces tsukubaensis NRRL18488 Contig510, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AJSZ01000615.1	Streptomyces tsukubaensis NRRL18488 Contig615, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AJSZ01000788.1	Streptomyces tsukubaensis NRRL18488 Contig788, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AJSZ01000843.1	Streptomyces tsukubaensis NRRL18488 Contig843, whole genome shotgun sequence	2577	2577	100%	0.0	97%
AGDC01001264.1	Streptomyces lysosuperificus ATCC 31396 contig327.4, whole genome shotgun sequence	2575	2575	100%	0.0	97%
AJGV01000235.1	Streptomyces auratus AGR0001 Scaffold16_1, whole genome shotgun sequence	2556	2556	100%	0.0	97%
AGDF01000186.1	Streptomyces sp. Wigar10 contig00190, whole genome shotgun sequence	2555	2555	100%	0.0	97%
NC_016114.1	Streptomyces flavogriseus ATCC 33331 chromosome, complete genome	2555	15296	100%	0.0	97%
AGSW01000013.1	Streptomyces sp. W007 contig00015, whole genome shotgun sequence	2549	2549	100%	0.0	97%
ADFC02000001.1	Streptomyces griseus XylebKG-1 scaffold_1_Cont1, whole genome shotgun sequence	2543	10172	100%	0.0	97%
NC_010572.1	Streptomyces griseus subsp. griseus NBRC 13350 chromosome, complete genome	2543	15263	100%	0.0	97%
ADFC02000002.1	Streptomyces griseus XylebKG-1 scaffold_1_Cont2, whole genome shotgun sequence	2538	5071	100%	0.0	97%
NC_018750.1	Streptomyces venezuelae ATCC 10712, complete genome	2538	8546	100%	0.0	100%
NC_017586.1	Streptomyces cattleya NRRL 8057 = DSM 46488 chromosome, complete genome	2538	15152	100%	0.0	97%
NC_016111.1	Streptomyces cattleya NRRL 8057, complete genome	2538	15114	100%	0.0	97%
NC_013929.1	Streptomyces scabiei 87.22 chromosome, complete genome	2538	15212	100%	0.0	97%
NC_015957.1	Streptomyces violaceusniger Tu 4113 chromosome, complete genome	2534	15136	100%	0.0	97%
NC_015953.1	Streptomyces sp. SirexAA-E chromosome, complete genome	2532	15191	100%	0.0	97%
ABYX01000136.1	Streptomyces roseosporus NRRL 11379 cont3.136, whole genome shotgun sequence	2503	2503	100%	0.0	96%
AJUO01000270.1	Streptomyces globisporus C-1027 Scaffold76_1, whole genome shotgun sequence	2486	2486	97%	0.0	97%
ADX801001041.1	Streptomyces sp. SA3_actF contig01366, whole genome shotgun sequence	2401	2401	87%	0.0	99%
AJTQ01000380.1	Streptomyces sulphureus L180 contig_380, whole genome shotgun sequence	2392	2392	100%	0.0	95%
NC_016109.1	Kitasatospora setae KM-6054, complete genome	2348	21128	100%	0.0	94%
AEYX01000010.1	Streptomyces griseoaurantiacus M045 Contig010, whole genome shotgun sequence	2265	2265	83%	0.0	99%
AFWY02000006.1	Amycolatopsis sp. ATCC 39116 Amy39116scaffold_2_Cont6, whole genome shotgun sequen	2146	2146	99%	0.0	92%
AFWY02000008.1	Amycolatopsis sp. ATCC 39116 Amy39116scaffold_2_Cont8, whole genome shotgun sequen	2146	2146	99%	0.0	92%

Figure A2.7: rDNA (16 S) from *Streptomyces coelicolor* A3(2), it will allow sequencing since it is house keeping among *Streptomyces* spp.