Spring 2009

Molecular Phylogenetic Analysis of Novel Spiroplasma Isolates

Haritha Lakshmi Jandhyam

MOLECULAR PHYLOGENETIC ANALYSIS OF NOVEL SPIROPLASMA ISOLATES

by

HARITHA LAKSHMI JANDHYAM

(Under the Direction of Laura B. Regassa)

ABSTRACT

Spiroplasmas are among the smallest self replicating organisms known. They are helical, motile descendents of Gram-positive bacteria that have evolved to occupy broad host ranges including plants, insects and crustaceans. Spiroplasmas are the causative agent of several economically important agricultural diseases, but most host-microbe interactions appear to be commensal. Given the ubiquitous nature of these microbes, a methodical approach that focused on serologically distinct isolates from a single host family was used as an initial step in understanding spiroplasma diversity and distribution. Tabanid-associated spiroplasmas represent the most thoroughly studied group to date, so this project examined a spiroplasma field isolate collection (>200 isolates) that was obtained from tabanid flies in Costa Rica, Ecuador, Australia, and the United States. The 16S rRNA, 23S rRNA, and *rpoB* genes and the 16S-23S rRNA spacer region were successfully used to establish evolutionary relationships of the closely related spiroplasma isolates. Phylogenetic analyses and non-genetic character mapping indicated that all study isolates belonged to the Apis clade; arginine hydrolysis was a strong indicator of evolutionary relatedness; surface serology and phylogenetic placement were congruent; and neither host specificity nor geographical ranges were strict.

INDEX WORDS: Mollicutes, Spiroplasma, Phylogenetics, Surface serology, Tabanids, Biodiversity, Host specificity,

MOLECULAR PHYLOGENETIC ANALYSIS OF NOVEL SPIROPLASMA ISOLATES

by

HARITHA LAKSHMI JANDHYAM

B.V.Sc & AH, Acharya N.G. Ranga Agricultural University, India, 2004

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2009

MOLECULAR PHYLOGENETIC ANALYSIS OF NOVEL SPIROPLASMA ISOLATES

by

HARITHA LAKSHMI JANDHYAM

Major Professor: Laura B. Regassa

Committee: Lorenza Beati
Oscar J. Pung

Electronic Version Approved:

May 2009

DEDICATION

I consider myself very lucky to have Krishnan as my loving, understanding and caring husband. I would like to thank him for his patience and support. I would like to dedicate my research work to my husband Krishnan, to my baby girl and to my parents Ramanaiah and Tulasamma. Thanks to my sisters, brother, aunt and uncle for all their loving care and support.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Laura B. Regassa, for giving me the precious opportunity to work with her and for her constant support, guidance and encouragement which allowed me to venture into new areas and grow as a researcher. I would like to thank Dr. Lorenza Beati for her expertise in phylogenetics and for the time she spent on my research work. She is very understanding, patient and overall a great person to work with. I would like to thank Dr. Oscar J. Pung, for spending his valuable time on my research project and for guiding me during the course of my graduate program. I would also like to thank Dr. Frank E. French for generously providing his strain collection and serological analyses, as well as his on-going support and expertise. In addition, I thank Ryan Bates for his technical assistance in lab and Tammie Young for generating the 16S rDNA sequences.

Funding and facilities for this research were provided by the Department of Biology, Georgia Southern University (Graduate Student Research Fund, Faculty Research Grant), the National Science Foundation (DEB-0481430, PI: Regassa), and the Southeastern Branch of the American Society for Microbiology.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Spiroplasmas (Procaryotae: *Mollicutes: Entomoplasmatales: Spiroplasmataceae*) are helical and motile Gram positive prokaryotes that are associated with insects and plants. The genus *Spiroplasma* was discovered recently and has drawn much attention due to its reductive evolution; characterized by small genome sizes (780 – 2,200 kb), low GC content (24% – 31%), and the lack of a cell wall (Carle et al. 1995, Williamson et al. 1997). Spiroplasmas may be one of the most abundant groups of microbes found on earth, due to their close association with a wide variety of organisms, ticks, plants and crustaceans, including tabanid flies (Hackett and Clark 1989, Hackett et al. 1992, Tully et al. 1987; Williamson et al. 1989; Williamson et al. 1998; Wang et al. 2004, 2005; Nunan et al. 2004). Although insect and plant hosts play a major role in maintenance and dispersal of these spiroplasmas, plant surfaces appear to be more important than plant phloem (Davis R E 1978; McCoy et al. 1979; Clark et al. 1987). Most spiroplasma host interactions appear to be commensal but there have been accounts of pathogenicity resulting in severe losses in agriculture and aquaculture industries (reviewed in Regassa and Gasparich 2006).

Based on 16S rDNA analysis, the class *Mollicutes* is a terminus in the evolution of gram-positive bacteria that arose from the low G+C content clostridial lineage of the Eubacteria (Woese et al. 1980, Weisburg et al. 1989, Johansson and Petterson 2002, Gasparich et al. 2004). *Mollicutes* are divided into two major branches, the *Acholeplasma-Anaeroplasma* branch and the *Mycoplasmatales-Entomoplasmatales* branch. The *Acholeplasma-Anaeroplasma* lineage consists of *Asteroleplasma*,

Anaeroplasma, and *Acholeplasma* species and the *Mycoplasmatales-Entomoplasmatales* lineage consists of *Spiroplasma*, *Entomoplasma*, *Mesoplasma*, *Mycoplasma*, and *Ureaplasma*. The *Mycoplasma-Ureaplasma* lineage derived from a spiroplasmal ancestor, since the complex cytoskeletal structure and larger genome size of spiroplasma appear to be primitive (Gasparich et al. 2004).

Host specificity

The level of host specificity appears to vary for spiroplasmas. A high degree of host specificity has been described for *S. poulsonii*, *S. kunkelli* and *S. leptinotarsae*. *S. poulsonii* causes sex-ratio abnormalities in the neotropical fruit fly *Drosophila willistoni* (Williamson et al. 1979, 1999). *S. kunkelli* has an obligate life cycle pattern involving both plant phloem and the leafhopper *Dalbulus maidisli* (Davis R E et al. 1978, Whitcomb et al. 1986, Hackett and Clark 1989, Ebbert and Nault 1994). Strict host specificity is also seen in *S. leptinotarsae*; it was isolated from the Colorado potato beetle but is unable to infect other insects (Hackett and Clark 1989, Hackett et al. 1996). Similarly, *S. floricola* can multiply in some beetles but not in the Colorado potato beetles (Clark and Whitcomb 1984), suggesting that certain host defense mechanisms exist that limit host switching. Tabanid-associated spiroplasmas are the most widely studied group of spiroplasmas; they do not appear to have strict host specificity at the species level (Clark et al. 1984, Regassa and Gasparich 2006). Plant surfaces are more important in dispersal and transmission of tabanid-associated spiroplasmas than plant phloem, indicating a horizontal mode of transmission (Davis R E 1978; McCoy et al. 1979; Clark et al. 1987). Although flowers do not play a major role in transmission of insect spiroplasmas, they can serve as storage and

exchange surfaces for spiroplasmas (Clark et al. 1984). Apart from insects and plant hosts, spiroplasmas are also able to infect crustaceans, ticks and higher order vertebrates (Tully et al. 1987, Williamson et al. 1989, 1998, Nunan et al. 2004, Bi et al. 2008); the level of host specificity in these organisms is unknown.

Pathogenicity

Spiroplasma interactions with their insect or plant hosts are usually classified as commensal, but some cases of pathogenicity or mutualism have been described. Spiroplasmas generally present in the salivary glands, crop, gut and hemolymph of tabanid hosts. Arthropod-borne spiroplasmas usually attach to the epithelial cells of the host midgut in a commensal relationship with their host (Hackett and Clark 1989). In cases where the spiroplasmas have adapted to invade tissues beyond the midgut, insect diseases can occur. For example, hemolymph invasion by *S. apis* and *S. melliferum* is linked to honey bee mortality (Clark T B 1978, Mouches et al. 1983); and transovarial transmission in *Drosophila* species cause sex ratio disorders limiting the male progeny (Williamson et al. 1979, 1999). The three major phytopathogenic spiroplasmas are *S. citri* that was discovered in 1973 and causes citrus stubborn disease and brittle root disease (Calavan and Bove 1989, Saglio et al. 1973); *S. kunkelii* discovered in 1975 that causes cornstunt disease; and *S. phoeniceum* discovered in 1986 that causes periwinkle disease. Generally, phytopathogenic spiroplasmas are restricted to plant sap and are transmitted by sap feeding insect vectors (Bove 1997). More recently discovered crustacean spiroplasmas cause a tremor disease that is responsible for heavy economic loses in the aquaculture industry (Bi et al. 2008). Although spiroplasmas are able to replicate at 37ºC, there is only one

report of a naturally-acquired human infection (Lorenz et al. 2002) in addition to experimental animal infections (Kirkhhoff et al. 1981, Tully et al. 1982, Megraud et al. 1983).

Species Designation

The minimal standards for spiroplasma species descriptions were recently revised to incorporate a polyphasic approach (Brown et al. 2007)**.** The standards focus on morphological, serological, biochemical and molecular properties. Some of the main descriptive characters are explained below.

Serology. The classification of spiroplasmas into provisional taxa based on serological and genomic properties began with the proposal of five groups (Junca et al. 1980). The accepted standard of 70% or greater DNA-DNA homology for a species worked well for the initial groups of spiroplasmas; and surface serology was eventually chosen as a surrogate for DNA-DNA hybridization (International Committee on Systematic Bacteriology [ICSB] 1995). Some groups were further divided into subgroups, with 30-70% homology and various degrees of cross-reactive serology (ICSB 1995). To date, a total of 38 groups have been recognized (Bove et al. 1983; Gasparich et al. 1993; Williamson et al. 1998; Nunan et al. 2005, Whitcomb et al. 2007); with groups I, VIII and XVI further divided into nine, three and three subgroups respectively. Reciprocal serological deformation tests are routinely used for serogroup placement (Tully 1983, Whitcomb and Hackett 1987).

Structure/Morphology. Spiroplasmas are 0.2 to 0.6μM in diameter with a variable length of 1-8μM (Trachtenberg et al. 2004). Dark field microscopy is used to observe the motility and helicity of spiroplasmas, whereas electron microscopy is used to determine

the lack of a cell wall. This lack of a cell wall makes all spiroplasmas resistant to penicillin. Morphologically, spiroplasma cells vary. Serogroup VIII spiroplasmas are short and thin; while spiroplasmas in other serogroups are generally characterized by long cells. All spiroplasmas are motile and helical but helicity varies among serogroups. For example, *S. ixodetis* exhibits tightly coiled helicity as compared to the other spiroplasmas (Tully et al. 1995). The spiroplasmas exhibit flexing and rotational movements along the longitudinal axis (Moulder et al. 2002).

Isolation. Isolation of pure spiroplasma cultures from the host relies on a triple dilution cloning method (Brown et al. 2007). Initially, spiroplasmas are separated from other microbes based on their ability to pass through a 220nm filter. The culture obtained after initial filtration is then subjected to a dilution cloning technique three times to obtain a clonal isolate (Tully 1983, Whitcomb and Hackett 1987).

Media and colony morphology. Spiroplasmas require serum components in the media for their growth. Generally, M1D media is used to culture the spiroplasma (Whitcomb 1983). The M1D contains insect cell extracts, fetal bovine serum, glucose, peptones, tryptone and amino acids. Spiroplasmas usually are not grown on solid media because motility leads to diffuse and/or satellite colonies (Tully and Whitcomb 1991, Whitcomb and Tully 1984).

Growth characteristics. All spiroplasmas are facultative anaerobes. They reproduce by binary fission with doubling times of 0.7 to 36.7 hours in M1D media (Williamson et al. 1998). The optimal growth temperatures vary. Some of the group I strains grow only at 25ºC and 30ºC; whereas almost all of the tabanid-associated spiroplasmas grow optimally at 30-32ºC, with a few exceptions at 25ºC and 37ºC

(Konai et al. 1996a; Helias et al. 1998).

Biochemical properties. Traditional biochemical characters are of limited utility for spiroplasma descriptions. All spiroplasmas studied to date are able to ferment glucose and none are able to hydrolyze urea (reviewed in Brown et al. 2007). Arginine hydrolysis varies in spiroplasmas; Group VIII strains catabolize arginine but most others do not (Hackett et al., 1996a; Moulder et al. 2002). Some of the spiroplasmas do not require sterol for growth (Rose et al., 1993), but this character and the ability to stringently test for it have been problematic resulting in its removal from the prescribed tests for new species descriptions (Brown et al. 2007). Finally, as discussed above, all spiroplasmas exhibit resistance to penicillin (500U/ml) due to lack the lack of a cell wall.

Base composition. Spiroplasmas are A-T rich microbes with some of the smallest genomes known for free living organisms. Genome sizes in spiroplasmas vary from 780 kbp for *S. platyhelix* to 2220 kbp for *S. ixodetis* (Carle et al. 1995, Tully et al. 1995; Williamson et al. 1997). Generally G + C composition in spiroplasmas range between 24-31% as determined by the melting temperature method (ICSB 1995). In spiroplasmas, one of the stop codons (UGA) encodes for tryptophan; Gasparich and collegues (2004) suggested that this characteristic may be synapomorphic for the *Mycoplasmatales-Entomoplasmatales* clade, which contains the *Spiroplasma* species.

Evolutionary Relationships

Recent phylogenetic reconstructions of the genus *Spiroplasma* examined the evolutionary relationship of the 34 described *Spiroplasma* species/serogroups (Table 1), along with the other members of the order Entomoplasmatales, based on 16S rRNA gene sequences (Gasparich et al. 2004). The order Entomoplasmatales contains four major clades: the Mycoides-Entomoplasmataceae Clade, the Apis Clade, the Citri-Chrysopicola-Mirum Clade, and the Ixodetis Clade (Figure 1). These four clades include *Spiroplasma, Mycoplasma, Entomoplasma* and *Mesoplasma.*

The Ixodetis Clade represents a single strain, *S. ixodetis* (serogroup VI), that is basal to the spiroplasma evolutionary tree (Gasparich et al. 2004). The *S. ixodetis* exhibits the largest genome size (2220 kbp) in the genus *Spiroplasma* with a peculiar, tightly coiled helicity (Tully et al. 1995).

The Citri-Chrysopicola-Mirum Clade represents three major clades: (1) The Citri-Poulsonii clade, (2) The Chrysopicola-Syrphidicola-TAAS-1clade, and (3) The Mirum clade. The Citri-Poulsonii clade contains serogroup I and II strains. Serogroup I spiroplasmas are further divided into nine subgroups consisting of pathogenic organisms with diverse host relationships like honey bees, plant surfaces, leaf hoppers and shrimp. Some of the important agricultural diseases caused by this group of spiroplasmas are citrus stubborn disease (*S. citri*), corn stunt disease (*S. kunkelli*), honey bee mortality (*S. melliferum*) and the recently described shrimp aquacultural diseases (*S. penaei*). The Chrysopicola-Syrphidicola-TAAS-1 clade contains serogroup VIII strains that represent the smallest spiroplasmas with a G+C content of 28-31mol% and the ability to catabolyze arginine. The clade contains the three

subgroup type strains *S. syrphidicola* (VIII-1), *S. chrysopicola* (VIII-2) and TAAS 1(VIII-3). 16S-23S rDNA analysis also placed the strain BARC 2649 in group VIII as a new subgroup (Regassa et al. 2004). The Mirum clade contains *S. mirum*, a tick spiroplasma basal to the Citri-Chrysopicola-Mirum clade.

The Apis Clade represents the major clade among the four spiroplasma clades and most of the tabanid-associated spiroplasmas fall into this clade. The Apis clade consists of two basal clades closely related to the Mycoides-Entomoplasmataceae clade. The Lampyridicola-Leptinotarsae clade represents serogroups XIX and XX isolated from fire flies and Colorado potato beetles. The clade Sabaudiense–Alleghenense-TIUS-1 represents serogroups XIII, XXVI and XXIX isolated from different groups of insects. Some of the important clades in the Apis clade include the Apis-Montanense clade that consists of sister species isolated from tabanid flies; *S. apis* (serogroup IV) and *S. montanense* (serogroup XXXI) are separated by a very small evolutionary distance. The Litorale-Turonicum-Corruscae clade represents serogroups XIV, XVII and XVIII that were also isolated from tabanids. The Helicoides-Gladitoris-BARC1901 clade represents serogroups XXIII, XXXII and XXXIV strains. Clade BIUS-1-W115 members isolated from flower surfaces, represent serogroups XXI and XXX respectively. The Floricola-Diabroticae clade contains members isolated from flower surfaces and beetles, and represents serogroups III and XII. *Spiroplasma diminutum*, isolated from a mosquito, is a sister group to the clade CB-1-Ar 1357 that represents serogroup XVI.

The Mycoides-Entomoplasmataceae Clade contains Mycoides and Entomoplasmataceae clades. The Mycoides clade contains *M. mycoides*, the

Mycoplasma type strain that infects ruminant animals and is of major veterinary importance. The Entomoplasmataceae clade has *Entomoplasma* and *Mesoplasma* species which do not form distinct clades and generally share a common habitat.

Biogeography

The biogeography of spiroplasmas is poorly understood. It would seem reasonable that host ranges would define spiroplasma biogeography, but this does not always clarify the issue, especially for generalists such as those associated with tabanid flies. Limited sampling has also been a major problem, with the vast majority of isolates coming from North America.

Despite these challenges, biogeographical distribution patterns for some spiroplasmas have begun to emerge. For spiroplasmas with high host specificity, occurrence in a particular geographical area depends on the availability of their host. For example, *S. kunkelii* is wide spread from the southern United States to Argentina where its host, the leaf hopper *Dalbulus maidisli*, occurs (Whitcomb et al. 1989, Bove 1997). Extensive sampling in North America has begun to address ranges for some generalists. *S. litorale* and *S. montanense* have very limited geographical regions restricted to Georgia and North Carolina (Whitcomb et al. 1997). Similarly, *S. tabanidicola* and *S. gladitoris* are limited to the south eastern United States. *S. chrysopicola* is present throughout the United States, whereas *S. helicoides* was collected only in Oklahoma and the southeastern United States. (Whitcomb et al. 1997, Regassa and Gasparich 2006). Recent analysis of tabanids from the Costa Rican highlands identified 13 independent serogroups, with four of these being geographic variants of species previously identified in the United States (Whitcomb et al. 2007).

Analysis of more spiroplasma isolates from geographically diverse areas will be needed to gain a more complete picture of spiroplasma biodiversity and ranges.

Study Objectives

The overall goal of this study was to examine the geographic diversification of spiroplasma isolates from three continents based on their genetic relatedness and place these geographical isolates within a larger evolutionary context. Specifically, I resolved evolutionary relationships by generating 16S-23S rDNA spacer region, *rpoB* and 23S rRNA gene sequences for use in combination with the 16S rDNA sequence for each species. Overall, this study expanded the pre-existing data set of eight tabanid-associated isolates from North America and Europe (Gasparich et al. 2004) to include new isolates from North America, Meso-America, South America and Australia.

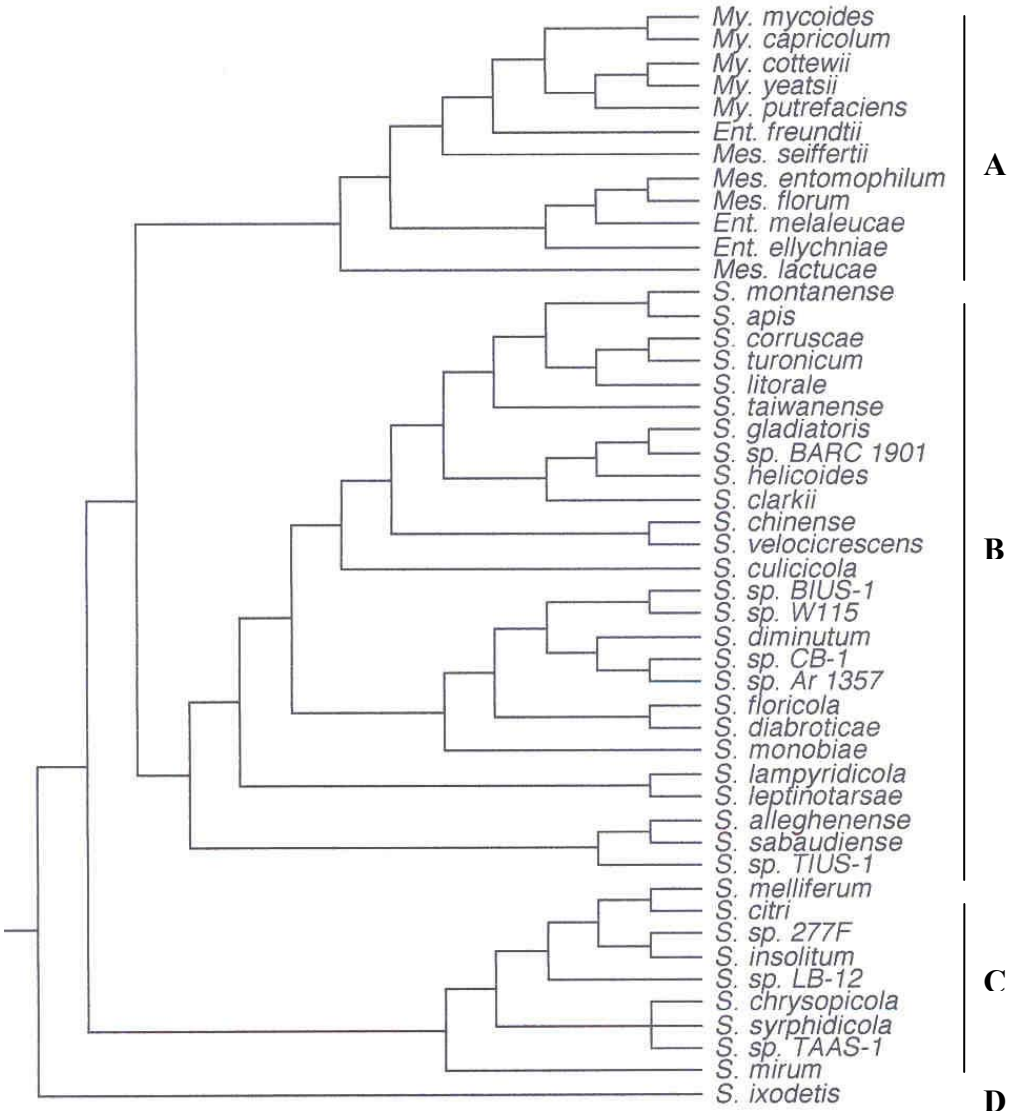

Figure 1. Majority Rule Maximum Parsimony tree based on 16S rRNA gene sequence analysis, showing position of major clades of the Spiroplasma-Entomoplamataceae-Mycoides (SEM) clade:
A) Mycoides-Entomoplasmataceae Clade, B) Apis Clade C) Citri-Chrysopicola-Mirum Clade and D) Ixodetis Clade. Specific parameters and GenBank accession numbers previously provided by Gasparich and co-workers (Gasparich et al. 2004) Figure modified from Regassa and Gasparich (2006).

Table1. *Spiroplasma* species and/or representative strains by serogroup[1]

Serogroup[2]	Binomial Name	Host(s)	Disease
I-1	*Spiroplasma citri*	Dicots, leafhoppers	Citrus stubborn
I-2	*S. melliferum*	Honey bees	Honeybee spiroplasmosis
I-3	*S. kunkelii*	Maize, leafhoppers	Corn stunt
I-4	*S.* sp.	Rabbit ticks	NK[3]
I-5	*S.* sp.	Green leaf bugs	NK
I-6	*S. insolitum*	*Eristalis* flies, flowers	NK
I-7	*S.* sp.	Coconut palms	NK
I-8	*S. phoeniceum*	*Catharanthus roseus*	Periwinkle disease
I-9	*S. penaei*	*Penaeus vannamei* shrimp	Shrimp disease
II	*S. poulsonii*	*Drosophila*	Sex ratio trait
III	*S. floricola*	Insects, flowers	Beetle "lethargy"
IV	*S. apis*	Bees, flowers	May disease
V	*S. mirum*	Rabbit ticks	Suckling mouse cataract disease
VI	*S. ixodetis*	*Ixodes pacificus* ticks	NK
VII	*S. monobiae*	*Monobia* wasps	NK
VIII-1	*S. syrphidicola*	*Eristalis arbustorum* flies	NK
VIII-2	*S. chrysopicola*	*Crysops* sp. Flies	NK
VIII-3	*S.* sp.	Horse fly	NK
IX	*S. clarkii*	*Cotinus* beetles	NK
X	*S. culicicola*	*Aedes* mosquitoes	NK
XI	*S. velocicrescens*	*Monobia* wasps	NK
XII	*S. diabroticae*	*Diabrotica undecimpunctata* beetles	NK
XIII	*S. sabaudiense*	*Aedes* mosquitoes	NK
XIV	*S. corruscae*	*Ellychnia corrusca* beetles, horse flies	NK
XV	*S.* sp.	Leafhopper	NK
XVI-1	*S. cantharicola*	Cantharid beetle	NK
XVI-2	*S.* sp.	Cantharid beetle	NK
XVI-3	*S.* sp.	Mosquito	NK
XVII	*S. turonicum*	Horse fly	NK
XVIII	*S. litorale*	*Tabanus nigrovittatus*	NK
XIX	*S. lampyridicola*	*Photuris pennsylvanicus* beetles	NK
XX	*S. leptinotarsae*	*Leptinotarsa decemlineata*	NK
XXI	*S.* sp.	*Prunus sp.* flowers	NK
XXII	*S. taiwanense*	*Culex tritaeniorhynchus*	NK
XXIII	*S. gladiatoris*	*Tabanus gladiator*	NK
XXIV	*S. chinense*	*Calystegia hederaceae*	NK
XXV	*S. diminutum*	*Culex* mosquito	NK
XXVI	*S. alleghenense*	Scorpionfly	NK
XXVII	*S. lineolae*	Horse fly	NK
XXVIII	*S. platyhelix*	Dragonfly	NK
XXIX	*S.* sp.	Tiphiid wasp	NK
XXX	*S.* sp.	Flower surface	NK
XXXI	*S. montanense*	Horse fly	NK
XXXII	*S. helicoides*	Horse fly	NK
XXXIII	*S. tabanidicola*	Horse fly	NK
XXXIV	*S.* sp.	Horse fly	NK
Ungrouped[4]	*S. atrichopogonis*	Biting midge	NK
XXXV	*S.* sp. BARC 4886	*Poeciloderas quadripunctatus*	NK
XXXVI	*S.* sp. BARC 4900	*Poeciloderas quadripunctatus*	NK
XXXVII	*S.* sp. BARC 4908	*Poeciloderas quadripunctatus*	NK
XXXVIII	*S.* sp. GSU5450	*Poeciloderas quadripunctatus*	NK

[1]Table modified from Regassa and Gasparich 2006.
[2]Serogroups are designated by Roman numerals; subgroups are indicated by hyphenated numbers.
[3]NK, none known
[4]No group number yet assigned to this serologically distinct species.

CHAPTER 2

MATERIALS AND METHODS

Study strains and growth conditions. A diverse collection of 211 spiroplasma field isolates were collected from female tabanid flies at seven locations in Australia, five in Costa Rica, one in Ecuador, and numerous sites in the Southern and Eastern United States (Regassa & Gasparich 2006). The spiroplasmas were isolated and pure cultures prepared using a dilution cloning technique (Tully 1983, Whitcomb and Hackett 1987). All strains were evaluated using a serological deformation test (Williamson et al. 1978, 1979 & 1998). The collection represents 29 independent serogroups, with 24 of these being novel groups. This study focused on representatives from each novel serogroup; strains used in this study are listed in Table 2. Spiroplasma strains were routinely grown in M1D media (Whitcomb et al. 1983), and cultures were observed by dark field microscopy to verify growth/morphology.

Serological validation. All strains used in this study were serologically validated using a deformation test (Williamson et al. 1978, 1979 &1998). The deformation test reaction mixture consists of antigen and specific polyclonal antiserum in 1:10 serial dilutions. The reaction mixtures that exhibit $\geq$50% cell deformation (e.g. clumping, blebs) are considered positive for antigenic reactivity against a known specific antibody (Williamson et al. 1978, 1983, 1998). An isolate reacting with specific antiserum at a concentration $\leq$1:320 was considered to be a validated member of that serogroup.

DNA isolation. DNA was extracted from serologically validated cultures using the Qiagen DNeasy Kit (Qiagen Inc., Valencia, CA).

PCR amplification and sequencing of 16S rRNA, 23S rRNA, and *rpoB* genes and the 16S-23S ITS region. PCR primers were designed for each region (Table 3). The 16S rRNA gene was PCR amplified using primers previously described by Fukatsu and Nikoh (1998) with minor modifications. The 16S-23S intergenic spacer (ITS) region was PCR amplified using primers previously described by Regassa et al. (2004); the primers have homology to the 16S and 23S rRNA genes. The 23S rRNA and *rpoB* (RNA polymerase beta subunit) primers were designed by aligning the sequences of closely related *Spiroplasma* species from the Apis clade that were available in Genbank, with the exception of 23SF2, which was adapted with minor modifications (Bi et al. 2008). The 50µl PCR mixture used to amplify the target genes/regions contained 100 ng of genomic DNA, 100 pmol of each primer, 0.2 mmol/L dNTPs, and 2.5U taq polymerase. The magnesium chloride concentration varied for each gene (Table 3). Amplification cycle were as follows: denaturation at 94°C for 5 min; 35 cycles at 94°C for 45 sec, annealing at 45-50°C for 45-60 sec, extension at 72°C for 1-2 min; and a final extension at 70-72°C for 1-5 min (Table 3). The PCR products generated were separated on 1.8-2% agarose gels, excised and purified using the Qiagen Gel Extraction Kit (Qiagen Inc. Valencia, CA). The PCR primers were also used to sequence the PCR-amplified fragments at the Oklahoma State University DNA sequencing facility (Stillwater, OK). Additional internal primers were designed for the *rpoB*, 16S rRNA and 23S rRNA gene sequences (Table 3).

DNA sequences. Sequences were assembled and aligned using Sequencher Software (Gene Codes Corp. Ann Arbor MI), MacClade software (Maddison and Maddison 1992) and manual alignment. All regions were double-strand sequenced, with any discrepancies

between the two strands being resolved by additional sequencing reactions. All sequences were deposited in Genbank (Table 4).

Phylogenetic analysis. Phylogenetic trees were constructed using maximum parsimony, maximum likelihood and Bayesian analysis. Phylogenetic analyses were carried out using PAUP 4.0 (Swofford, 2001) and support for the tree topologies was assessed by 100 bootstrap replicates. Bayesian analysis was performed in MRBAYES 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), using four chains with six nucleotide substitution rates and a gamma distribution of rate variation between sites. Bayesian analysis was performed for 1 million generations and posterior probability values were obtained by computing a 50% majority rule consensus of the trees generated by the search, with the exclusion of trees obtained before likelihood values stabilized. Additional *Mollicutes* species were used to clarify the placement of the novel isolates (Table 5).

Table 2. Apis Clade *Spiroplasmas* from *Tabanidae*: *Diptera* of Australia, Costa Rica, Ecuador and U.S.A

Serogroup representatives by country of origin	Number of serologically similar isolates[1]	Host
Australia		
Tab4C (*S. turonicum*)	11	*Haematopota pluvialis*
GSU5510	-	*Tabanus townsvilli*
GSU5603	21	*Scaptia lasiophthalma*
GSU5529	-	*Cydistomyia sp*
GSU5478	6	*Tabanus pallipenis*
GSU5490	5	*Tabanus parvicallosus*
GSU5508	11	*Tabanus particaecus*
Costa Rica		
TABS-2 (*S. helicoides*)	8	*Tabanus abactor*
GSU5441[2]	-	*Tabanus praeteritus*
TALS-2 (*S. lineolae*)	6	*Tabanus lineola*
BARC 4903	-	*Poeciloderas quadripunctatus*
TN-1 (*S. litorale*)	7	*Tabanus nigrovittatus*
GSU5353[2]	-	*Dichelacera princessa*
BARC 4689	3	*Tabanus lineola*
GSU5400	-	*Dichelacera princessa*
BARC 4908	4	*Poeciloderas quadripunctatus*
BARC 4906		*P. quinquevitattus*
GSU5373	3	*Tabanus occidentalis*
GSU5420	-	*Tabanus occidentalis*
BARC 4900	2	*Poeciloderas quadripunctatus*
BARC 4886	1	*Poeciloderas quadripunctatus*
GSU5450	2	*Poeciloderas quadripunctatus*
GSU5382	8	*Tabanus punctipleura*
GSU5360	4	*Dichelacera princessa*
GSU5363	1	*Tabanus occidentalis*
GSU5443	3	*Tabanus oculus*
GSU5446H	3	*Tabanus nebulosus*
GSU5366	2	*Tabanus secundus*
GSU5405	1	*Tabanus occidentalis*
Ecuador		
GSU5862	3	*Tabanus occidentalis*
GSU5861	1	*Dichelacera albitibialis*
GSU5867	2	*Tabanus occidentalis*
GSU5865	1	*Tabanus occidentalis*
GSU5858	1	*Phaeotabanus nigriflavus*
GSU5853	1	*Phaeotabanus prasiniventris*
U.S.A.		
TAUS-1 (*S. tabanidicola*)	93	*Tabanus sp.*
BARC 4689	4	*Tabanus lineola*
GSU4980[3]	3	*Tabanus lineola*
TALS-2 (*S. lineolae*)	11	*Tabanus lineola*

[1]Number of serologically similar isolates from country of origin (not including serologically related isolates from another country.
[2]Serogroup placement based on strong one-way deformation test (1:640).
[3]Tentative serogroup placement; no antiserum generated against GSU4980

Table 3. Conditions for PCR amplification of *Spiroplasma* genes/region.

Gene/region	Primer	Primer sequence 5'-3'	PCR	Sequencing	Amplicon	Anneal temperature	Extension	Final extension	[$MgCl_2$]
16S rRNA	16SF1	AGAGTTTGATCMTGGCTCAG	X		1500 bp	55°C 1 min	72°C 2 min	72°C 5 min	1.5 mM
	16R1	TACGGATACCTTGTTACGACTT	X						
	16SF2	GTGCCAGCAGCCGCGGTAATAC		X					
	16SR2	CGAGCTGACGACATCCATGCA		X					
	16SF3	TGCATGGTTGTCGTCAGCTCG		X					
	16SR3	GTATTACCGCGGCTGCTGGCAC		X					
	16SF4	AGCCATTGTAGCACGTGTGTAG		X					
	16SR4	GAGAGTCAAGCCGATGCATA		X					
	16SF5	ATACGTAGCCGAACTGAGAG		X					
ITS	ITSF	CGGTGAATACGTTCTCG	X	X	450 bp	46°C 45 sec	72°C 1 min	72°C 1 min	1.5 mM
	ITSR1	CAAGGCATTCACCATAC	X	X					
23S rRNA	23SF1	GTTCTTTGAAAACTGAATATTAG	X		1550 bp	45°C 1 min	72°C 2 min	72°C 5 min	2.2 mM
	23SR1	TTCGCTACCTTAGGACCGTTATAG	X						
	23SF2	GCTGGAAGTAAGCTTTGATCC		X					
	23SR2	TGCAGAGTTCCTTAGCTACAG		X					
	23SF3	GTARACACGAAACCAGGTGA		X					
	23SR3	ACTAGCTATCTCCAGGTTCG		X					
rpoB	rpoBF1	CAYTAYGGWAGAATYTGTCC	X	X	1250 bp	45°C 1min	72°C 2min	72°C 5min	2.0mM
	rpoBR1	ACYTGACCRATRTTCATCCG	X	X					
	rpoBF2	CTCWGGDGATGCHRTTGTWGC		X					
	rpoBR2	TTCCATWGAWGGWCCRCTDGC		X					

Table 4. New accession numbers for DNA sequences generated from *Spiroplasma* isolates/species and deposited as part of this study.

Strains (by country of origin)	Gene/Region			
	16S rRNA	23S rRNA	16S-23S spacer (ITS)	*rpoB*
Australia				
GSU5510	FJ821668	FJ824548	FJ845711	FJ845744
GSU5603	FJ821669	FJ824549	FJ845712	FJ845745
GSU5529	FJ821670	FJ824550	FJ845713	FJ845746
GSU5478	FJ821671	FJ824551	FJ845714	FJ845747
GSU5490	FJ821672	FJ824552	FJ845715	FJ845748
GSU5508	FJ821673	FJ824553	FJ845716	FJ845749
Costa Rica				
GSU5441	FJ821674	FJ824554	FJ845717	FJ845750
BARC 4903	FJ821675	FJ824555	FJ845718	FJ845751
GSU5353	FJ821676	FJ824556	FJ845719	FJ845752
GSU5400	FJ821677	FJ824557	FJ845720	FJ845753
BARC 4908	FJ821678	FJ824558	FJ845721	FJ845754
BARC 4906	FJ821679	FJ824559	FJ845722	FJ845755
GSU5373	FJ821680	FJ824560	FJ845723	FJ845756
GSU5420	FJ821681	FJ824561	FJ845724	FJ845757
BARC 4900	FJ821682	FJ824562	FJ845725	FJ845758
BARC 4886	FJ821683	FJ824563	FJ845726	FJ845759
GSU5450	FJ821684	FJ824564	FJ845727	FJ845760
GSU5382	FJ821685	FJ824565	FJ845728	FJ845761
GSU5360	FJ821686	FJ824566	FJ845729	FJ845762
GSU5363	FJ821687	FJ824567	FJ845730	FJ845763
GSU5443	FJ821688	FJ824568	FJ845731	FJ845764
GSU5446H	FJ821689	FJ824569	FJ845732	FJ845765
GSU5366	FJ821690	FJ824570	FJ845733	FJ845766
GSU5405	FJ821691	FJ824571	FJ845734	FJ845767
GSU5358	FJ821692	FJ824572	FJ845735	FJ845768
Ecuador				
GSU5862	FJ821693	FJ824573	FJ845736	FJ845769
GSU5861	FJ821694	FJ824574	FJ845737	FJ845770
GSU5867	FJ821695	FJ824575	FJ845738	FJ845771
GSU5865	FJ821696	FJ824576	FJ845739	FJ845772
GSU5858	FJ821697	FJ824577	FJ845740	FJ845773
GSU5853	FJ821698	FJ824578	FJ845741	FJ845774
U.S.A				
BARC 4689	FJ821699	FJ824579	FJ845742	FJ845775
GSU4980	FJ821700	FJ824580	FJ845743	FJ845776
S. tabanidicola (TAUS-1)	FJ821701	FJ824581	ND[1]	ND[1]

[1]Not done as part of this study. DNA sequence available from prior studies; see Table 5 for accession numbers.

Table 5. Genbank accession numbers for sequences from described *Mollicutes* species used in the phylogenetic analyses.

Strain	**16S rRNA**	**ITS**	***rpoB***	**23S rRNA**
S. citri G113-3X	M23942	DQ004934	AM285302	AM285316
S. chrysopicola DF-1	AY189127	AY549216	NA	NA
S. apis B 31	M23937	AY736030	NA	NA
S. helicoides TABS-2	AY189132	DQ004903	DQ310574	NA
S. montanense HYOS-1	AY189307	DQ004925	DQ313817	NA
S. turonicum Tab-4c	AY189310	DQ439669	DQ313822	NA
S. corruscae EC-1	AY189128	DQ004943	NA	NA
S. litorale TN-1	AY189306	NA	DQ313822	NA
S. lineolae TALS-2	DQ860100	DQ860100	DQ310570	NA
S. gladitoris TG-1	M24475	DQ004899	DQ310573	NA
S. tabanidicola TAUS-1	NA	DQ004931	DQ310571	NA
S. sp. W115	AY189317	NA	NA	NA
S. sp. BIUS-1	AY189319	NA	NA	NA
S. diminutum CUAS-1	AY189130	DQ004921	NA	NA
S. mirum SMCA	M24662	DQ439670	M24662	DQ917756
S. ixodites Y-32	M24477	NA	NA	NA
M. mycoides PG-1	BX293980	BX293980	BX293980	BX293980
M. pneumoniae 129	AF132740	NA	NA	NA
S. floricola 23-6	AY189131	AY974061	NA	NA

CHAPTER 3

RESULTS

Spiroplasma isolates. The study strains were obtained from tabanid flies in Australia, Costa Rica, Ecuador and the United States (Regassa & Gasparich 2006). The spiroplasmas were isolated and pure cultures prepared using a dilution cloning technique (Whitcomb & Hackett 1987). The collection represents 29 independent serogroups, with 24 of these being novel groups. A total of 33 novel isolates and 1 described species (*S. tabanidicola*) were examined in detail in this study (Table 2).

Target gene/region DNA sequencing. 16S rRNA, 23S rRNA, and *rpoB* gene sequences and the 16S-23S rDNA intergenic spacer (ITS) region were chosen for phylogenetic analyses. The feasibility of these regions to resolve relationships in closely related species/isolates was demonstrated in other *Mollicutes* phylogeny (Volokhov et al. 2007; Bi et al. 2008). Each gene was PCR amplified using specific primers and sequences were double strand verified.

Phylogenetic Analyses. Phylogenetic analyses were carried out using maximum parsimony, maximum likelihood and Bayesian analyses to overcome the limitations of each individual method. The maximum parsimony heuristic search values for all study genes and the concatenated trees are shown in Table 6. For maximum likelihood, Model Test 3.7 (Posada and Crandall 1998) suggested that the GTR+I+G (general time reversible plus proportion of invariant sites plus gamma distribution of rate parameters) model best described the data. The –InL values were as follows: 16S rRNA: 5929.8760; Internal Transcribed Spacer (ITS): 1678.7141; 23S rRNA: 5643.5080; *rpob*: 6413.1522;

concatenated 16S & 23S rRNA: 8069.5522, concatenated 16S rRNA, ITS & *rpob*: 11079.00035, and concatenated 16S rRNA, ITS, 23S rRNA, & *rpoB*: 18520.7990.

The choice of outgroups for each analysis was dependent upon the target sequences being analyzed. For example, 16S and 23S rRNA genes are highly conserved and easily aligned with distantly related strains like *M. pneumoniae* and *S. citri*. In contrast, fast evolving genes/regions like the ITS and *rpoB* were analyzed with more closely related strains as outgroups or in unrooted trees. A summary of outgroups are as follows: 16S rRNA: *M. pneumoniae*; ITS: *S. floricola* and *S. diminutum*; 23S rRNA: *S. citri* and *S. mirum*; *rpoB*: no outgroup. To further support the groupings we also analyzed combinations of highly conserved genes (16S and 23S rRNA); rapidly evolving targets along with the traditional 16S rRNA gene (16S rRNA, ITS and *rpoB*); and all targets sequences (16S rRNA, ITS, 23S rRNA and *rpoB*). *S. citri* and *S. mirum* served as the outgroups for all concatenated sequences except for the concatenated ITS, *rpoB*, and 16S rRNA analysis that had no outgroup. Additional type strains were included to demonstrate the placement of the novel spiroplasma isolates in relation to previously described species (Table 6).

The individual trees generated using the three analytical models are shown in Figures 2-8. The Bayesian analysis (Figures 2A, 3A, 4A, 5A, 6A, 7A, 8A) contains the full complement of all strains in each grouping with identical sequences. In addition, the Bayesian analysis shows the relative position of *S. citri* and *S. mirum* outgroups for comparative analysis among target genes. The consensus of all analyses suggested five distinct groupings (I-V), as indicated by the figure coloring (Figures 2-8).

Group I consist of *S. montanense* from the United States and *S. apis* from France, as well as novel isolates from Australia (GSU5508, GSU5603, GSU5529) and Ecuador (GSU5858). Group II consists of *S. turonicum* from France and two isolates from Australia (GSU5478, GSU5510). Group III contains *S. corruscae* and *S. litorale* from the United States, in addition to isolates from Costa Rica (GSU5443, GSU5353, GSU5405, GSU5382) and Ecuador (GSU5861, GSU5865, GSU5853). Group IV consists of *S. lineolae*, *S. tabanidicola* and *S. gladiatoris* as representative strains from the United States; and new isolates from Australia (GSU5490), Costa Rica (BARC 4886, BARC 4900, BARC 4903, GSU5363, GSU5366, GSU5400, GSU5360), Ecuador (GSU5862), and the United States (BARC 4689, GSU4980). Group V consists of *S. helicoides* from the United States as a representative strain, 7 isolates from Costa Rica (BARC 4906, BARC 4908, GSU5450, GSU5420, GSU5441, GSU5446H,GSU5373), and 1 isolate from Ecuador (GSU5867).

While the overall phylogenetic analyses strongly supported the groups described above, there were two apparent discrepancies involving groups II and III. Groups II and III are well discriminated in all analyses except for trees based on the 16S rRNA gene (Figure 2), where they could be considered a single group. The 16S rRNA analysis was done by including distantly related strains (e.g. *S. ixodetis*, *M. pneumoniae*) to verify the placement of the study strains. This may contribute to the lack of differentiation between the closely related strains of groups II and III. The deep branching position of groups II and III varies among analyses; it appears as a basal group in 23S rRNA individual analyses, conc. 16S & 23S rRNA and conc. 16S rRNA, 23S rRNA, *rpoB* and ITS region (Figures 4, 6, 8). The deep branching boot strap values (75-98), the maximum parsimony

heuristic search values (CI: 0.911, RI: 0.807 and RC: 0.736), and the posterior probability values (100) were consistently strong for the individual 23S rRNA gene analysis as compared to the other genes, thereby supporting the basal position of groups II & III. The trees obtained from concatenated sequences that included the 23S rRNA gene also strongly supported the basal position of groups II and III with high bootstrap and posterior probability values.

Resolution within groups varied, with the greatest support observed for analyses involving the more rapidly evolving targets, *rpoB* and ITS. This is most evident in groups III, IV and V (Figures 3, 5, 7, 8). When comparing these analyses, concatenated trees with *rpoB* (Figures 5, 7, 8) and the individual *rpoB* analysis (Figures 5) provided more resolution than the ITS analysis (Figure 3). This may be due to the lower number of parsimony informative characters for the ITS analysis (113 characters); the *rpoB* analysis had 365 parsimony informative characters. It should be noted that group III was well supported with high posterior probability values (95-100) only when using concatenation of 16S rRNA and *rpoB* genes and the ITS region (Figure 7).

Character mapping. Preliminary character mapping for the novel isolates was completed for non-genetic characters, including surface serology, arginine hydrolysis, growth profiles, biogeography, and host specificity.

Serogroups were reported for each novel strain in a separate study (Regassa and Gasparich 2006). In this study, duplicate representatives were examined for 8 serogroups (Table 2) to evaluate the correlation between serology and phylogeny. Among the 4 serogroup paired strains that have a full complement of gene sequences, phylogenetic analyses placed all strains from the same serogroup within a small cluster in the same

major phylogenetic group. The placement of serogroups within the five major groups was as follows: Group I (GSU5603/GSU5529), Group IV (BARC 4689/GSU5400), and Group V (BARC 4906/BARC 4908 and GSU5420/GSU5373). Complete analyses of the other serogroups await generation of addition DNA sequences for target genes in the type strains. However, preliminary analyses based solely on the 16S rRNA gene sequence showed that all serogroup members were maintained within the same major phylogenetic group.

Nearly all study strains were unable to hydrolize arginine. Strains GSU5508, GSU5603, GSU5529 and GSU5858 hydrolyzed arginine; and all of these isolates fell into Group I. Group I also contained the arginine-positive species *S. apis* and *S. montanense*.

Growth profiles for all novel study isolates were determined at temperatures ranging from 10ºC to 41ºC (LB. Regassa and R. Bates, unpublished). For all but one study strain, optimal growth was observed at 30-32ºC, with doubling times of 0.4±0.1 hr to 1.6±0.1 hr except for two strains that grew slower (average 3.5±0.4 hr). Eight strains had optimal growth ranges that included 25ºC or 37ºC. The examination of optimal growth doubling times in relation to the phylogenetic groups revealed some general trends (Figure 9). Group I strains grew most aggressively with doubling times ranging from 0.4 to 0.8 hours (average 0.6 hr). Groups II, III and IV had average doubling times of 1.0, 0.9 and 1.2 hours, respectively. The ranges were 0.8-1.1 hr for group II and 0.8-1.2 hr for group III. The average doubling time for group IV is a bit misleading because GSU5490 was unusual (doubling time 3.6 hours); if GSU5490 is excluded then the average doubling time is 0.9 hour over a range of 0.6-1.3. The group V strains had the

longest average doubling time at 1.5 hours, with individual doubling times ranging from 1.0 to 3.3 hours.

Preliminary biogeographical analyses of the novel isolates were based on the host country. Australian isolates fell into groups I, II, and IV; Costa Rica isolates into groups III-V; Ecuador isolates into groups I, III, IV and V; and United States isolates into group IV. Examination of the 16S rRNA and concatenated phylogenetic analyses overlayed with the biogeography (Figures 2A and 8A) helped to clarify larger trends. There was major overlap between North and South American isolates (groups III, IV, V); all Costa Rican highland isolates were confined to groups IV and V; and all but one Australian isolate (GSU5490) fell into groups I and II.

A total of 27 distinct hosts were observed for the 41 novel isolates and characterized species that fell into the 5 phylogenetic groups of the Apis Clade (Figure 10). For all groups except group 1, *Tabanus* sp. were the most common host. *Tabanus* sp. were the hosts for 47% of the isolates in groups II, IV and V; and 56% of them in group III. *Poeciloderas* sp. and *Dichelacera* sp. were also common. *Poeciloderas* sp. served as hosts for 17% and 33% of group IV and V isolates, respectively. *Dichelacera sp.* were hosts for 22% of the group III isolates and 17% of group IV isolates. Group I showed the greatest host diversity; each of the six isolates was from a different genus. There does not appear to be strict host specificity for the spiroplasmas at the host genus or species level. This was evident within the phylogenetic groups, as described above, and also within the eight serogroup pairs. Six of the 8 serogroup pairs had hosts from 2 different genuses.

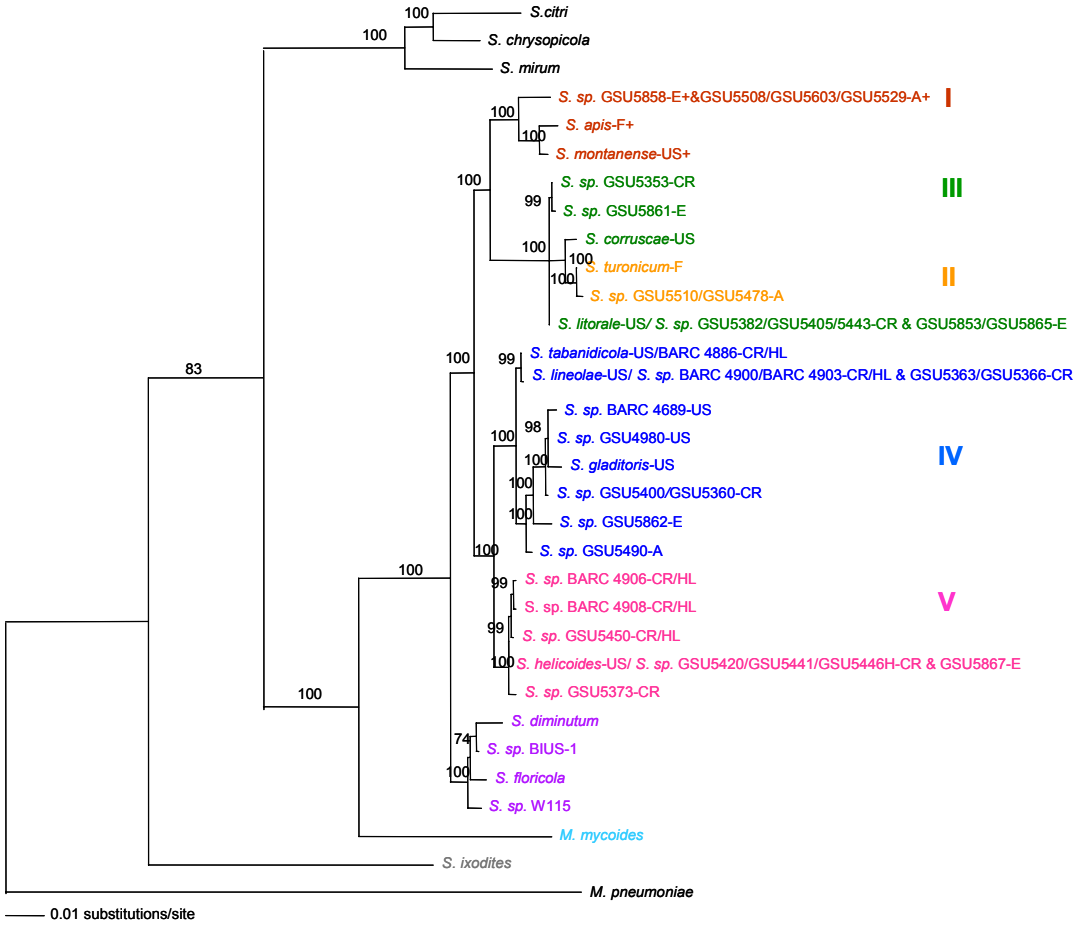

Figure 2A. Bayesian phylogenetic tree derived from partial 16S rDNA sequence analysis. The obtained 16S rDNA aligned matrix included a total of 1428 characters and *M. pneumoniae* was the out group. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

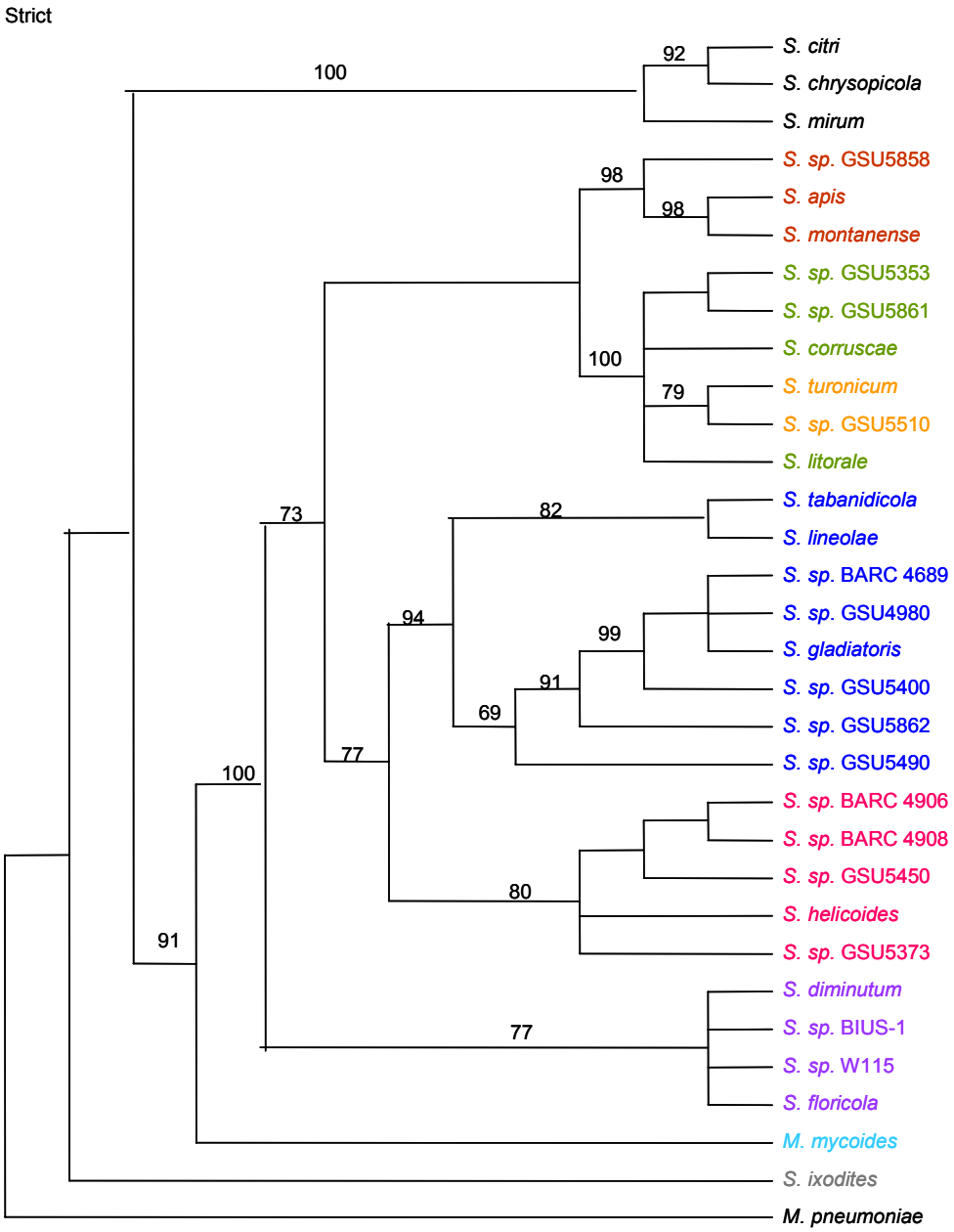

Figure 2B. Maximum Parsimony phylogenetic tree derived from partial 16S rDNA sequence analysis. The obtained 16S rDNA sequence aligned matrix included a total of 1428 characters and *M. pneumoniae* was the out group. Maximum Parsimony heuristic search in PAUP found 14 best trees (length=814, CI=0.693; RI=0.746; RC=0.517). 233 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates)

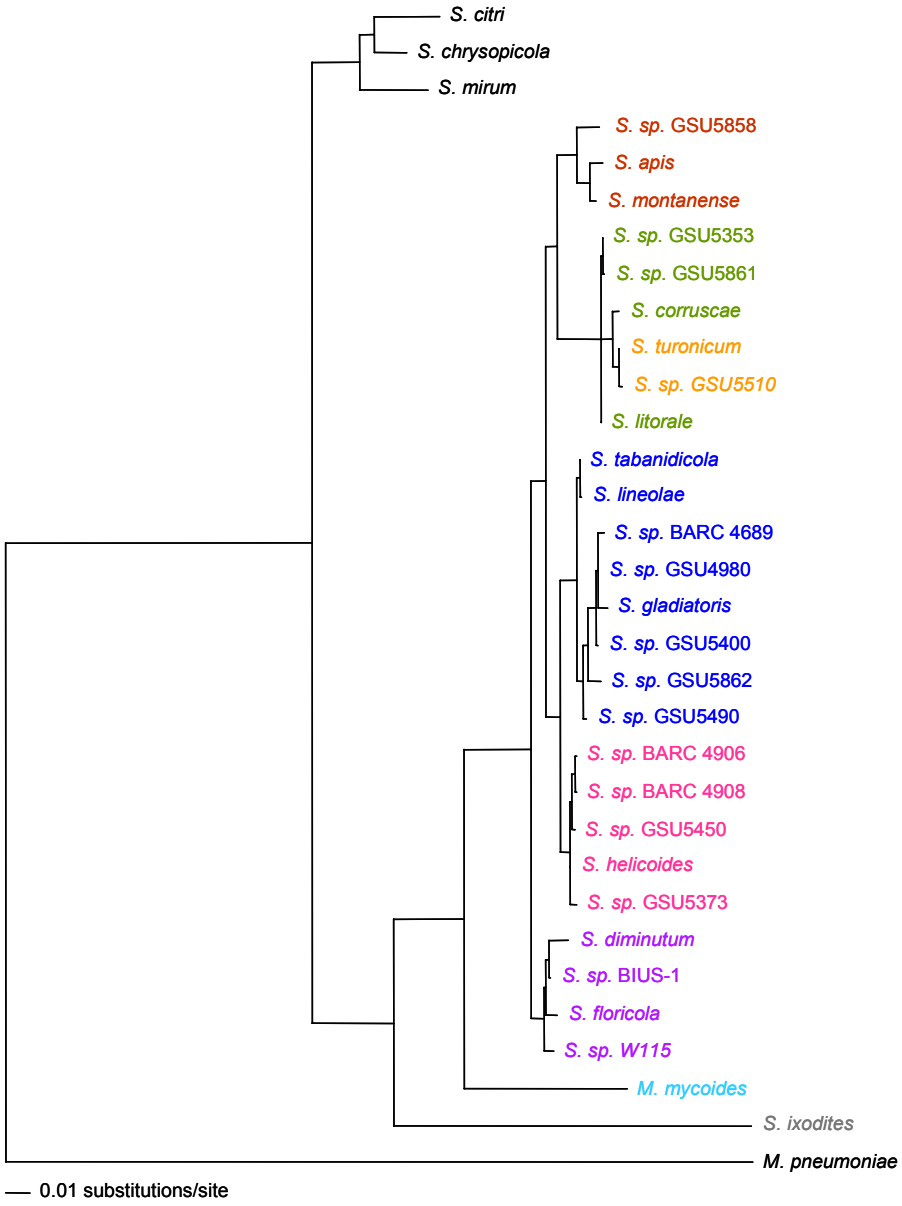

Figure 2C. Maximum Likelihood phylogenetic tree derived from partial 16S rDNA sequence analysis. The obtained 16S rDNA sequence aligned matrix included a total of 1428 characters and *M. pneumoniae* was the out group. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I; –InL value is 5929.8760.

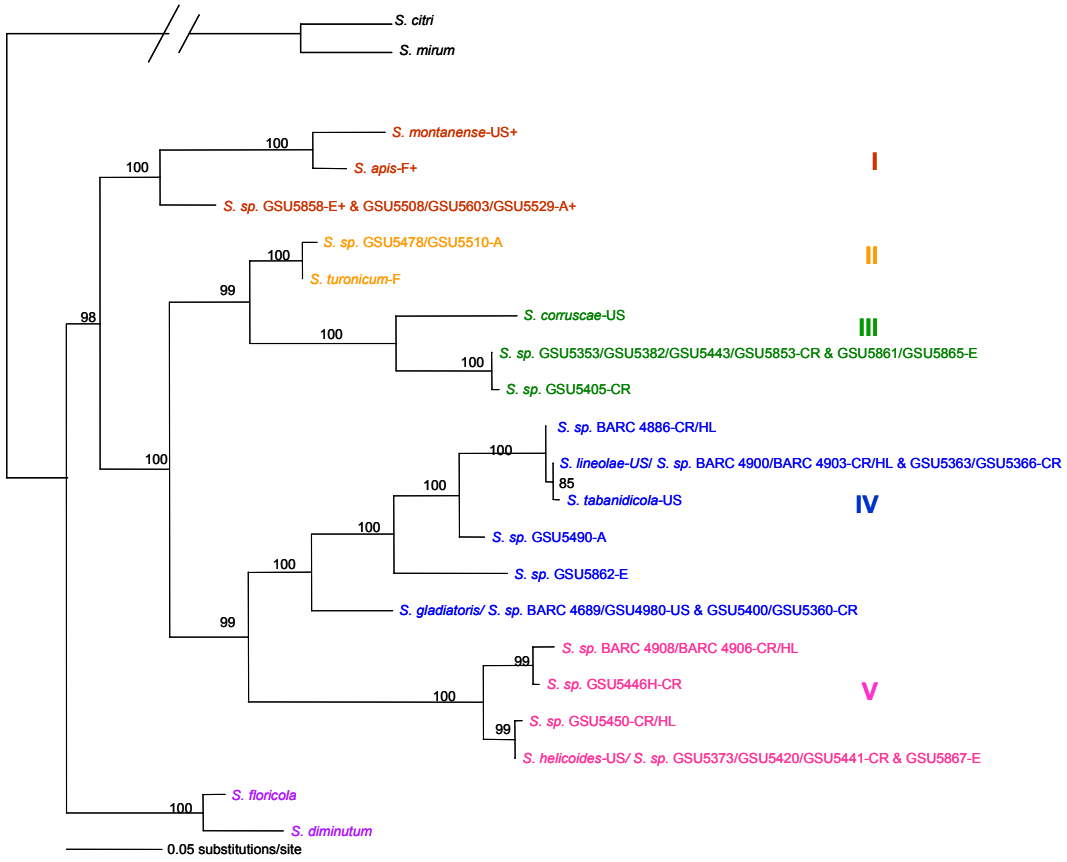

Figure 3A. Bayesian phylogenetic tree derived from partial 16S-23S rDNA sequence analysis. The obtained 16S-23S rDNA aligned matrix included a total of 307 characters, *S. floricola* and *S. diminutum* were the out groups. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican HighLands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

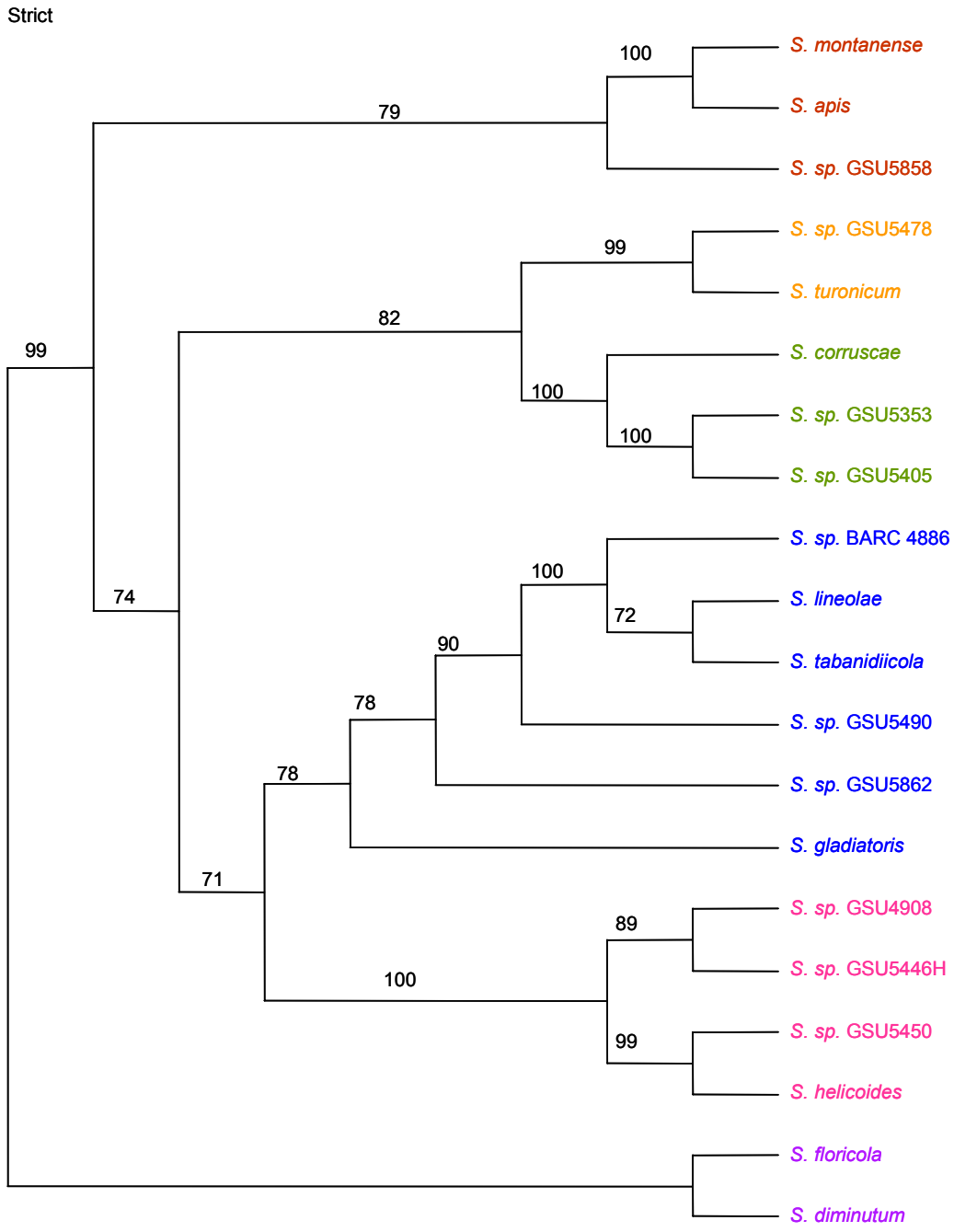

Figure 3B. Maximum Parsimony phylogenetic tree derived from partial 16S-23S rDNA sequence analysis. The obtained 16S-23S rDNA sequence aligned matrix included a total of 307 characters, *S. floricola* and *S. diminutum* were the out groups. Maximum parsimony heuristic search in PAUP found 1 best tree (length=281, CI=0.673; RI=0.795; RC=0.534). 113 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

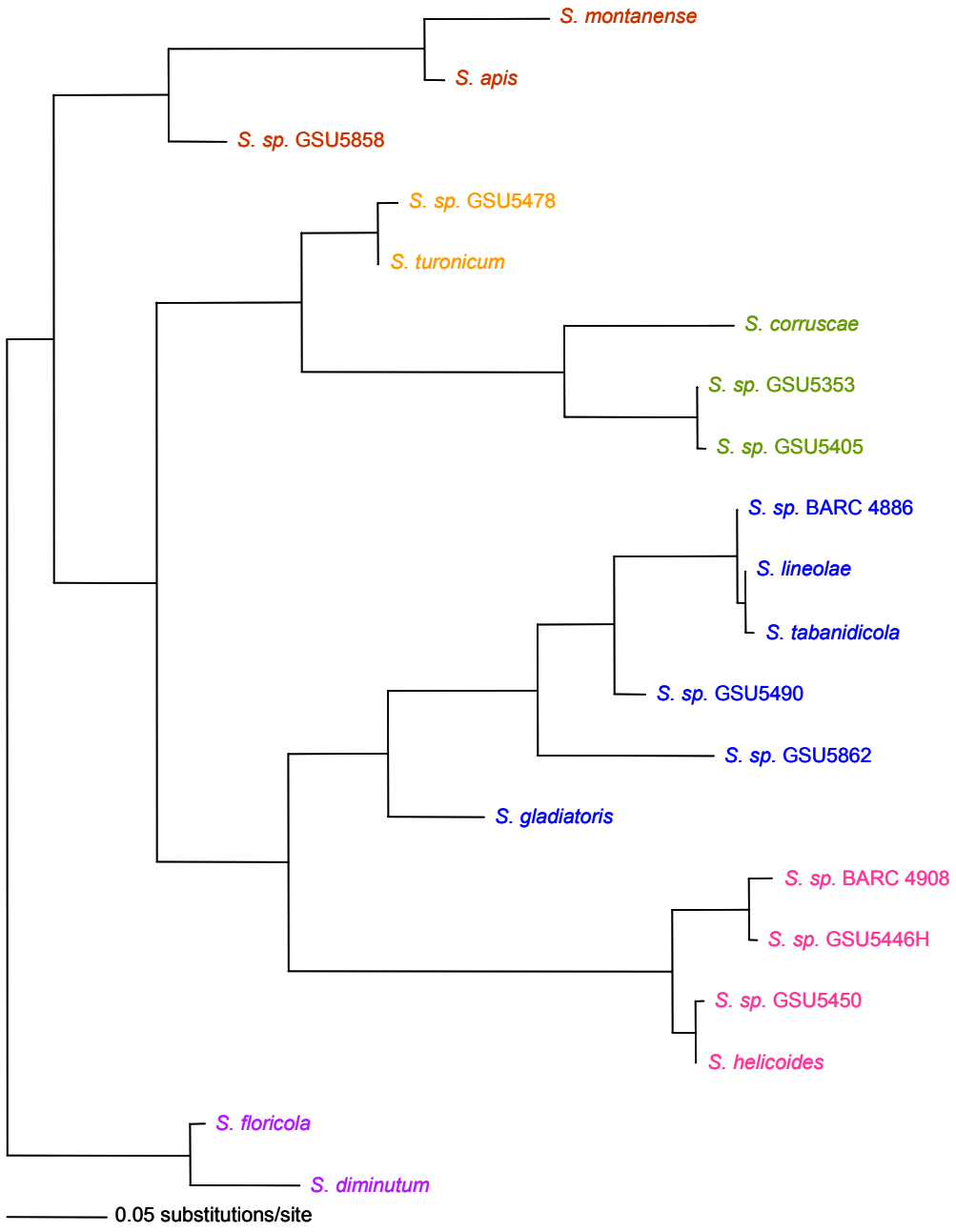

Figure 3C. Maximum Likelihood phylogenetic tree derived from partial 16S-23S rDNA sequence analysis. The obtained 16S-23S rDNA aligned matrix included a total of 307 characters, *S. floricola* and *S. diminutum* were the out groups. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 1678.7141.

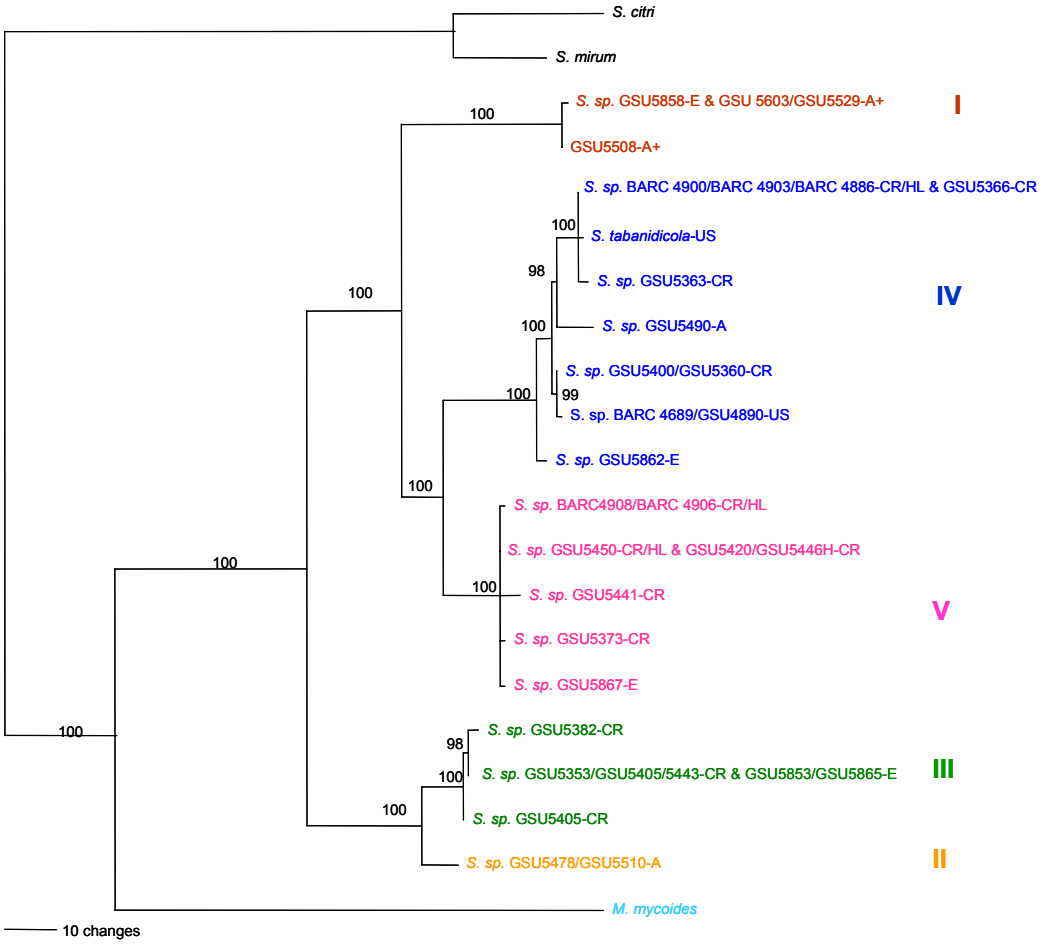

Figure 4A. Bayesian phylogenetic tree derived from partial 23S rDNA sequence analysis. The obtained 23S rDNA aligned matrix included a total of 1328 characters, *S. citri* and *S. mirum* were the out groups. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator.Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

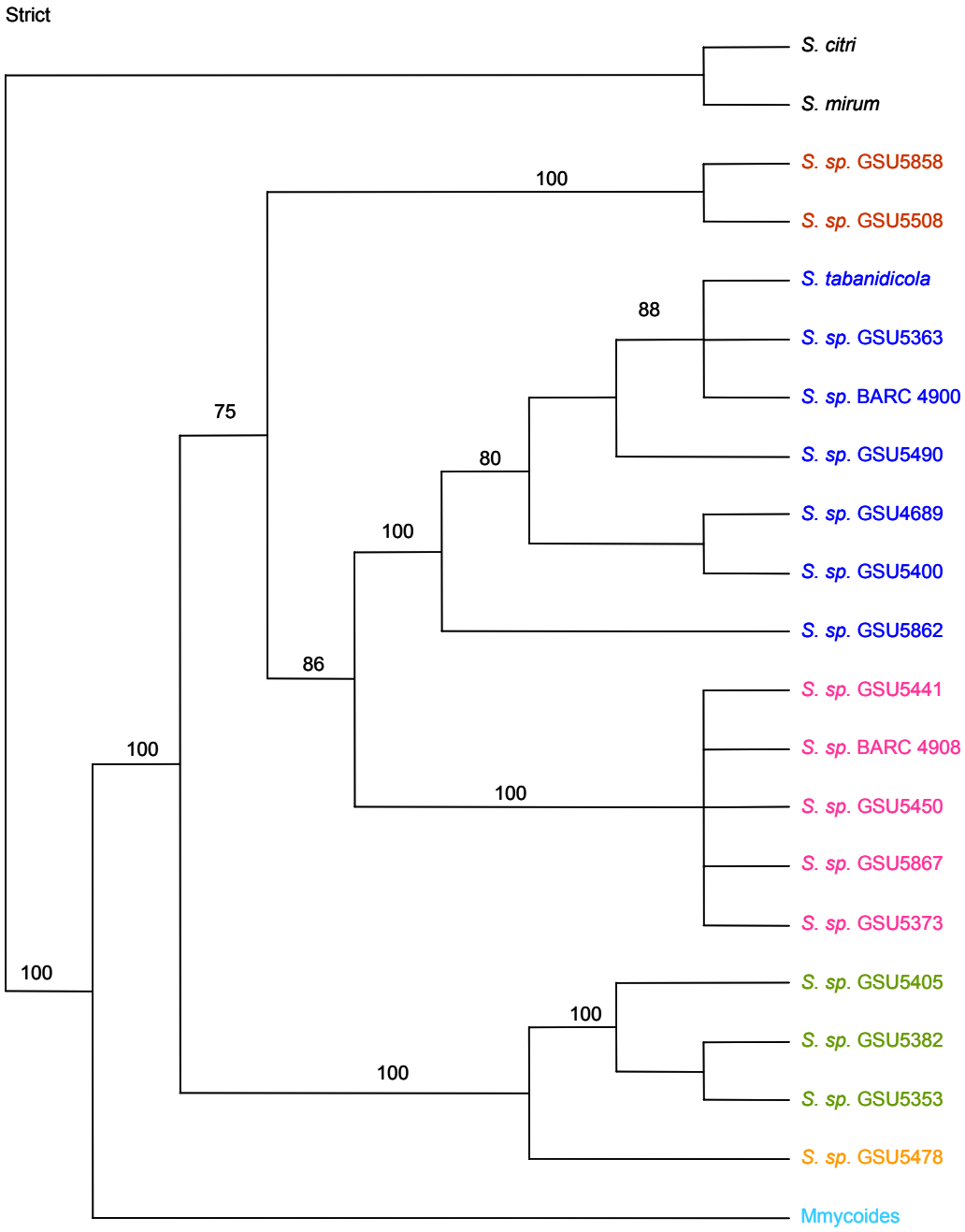

Figure 4B. Maximum Parsimony phylogenetic tree derived from partial 23S rDNA sequence analysis. The obtained 23S rDNA aligned matrix included a total of 1328 characters, *S. citri* and *S. mirum* were the out groups. Maximum Parsimony heuristic search in PAUP found 2 best trees (length=1214, CI=0.911; RI=0.807; RC=0.736). 182 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

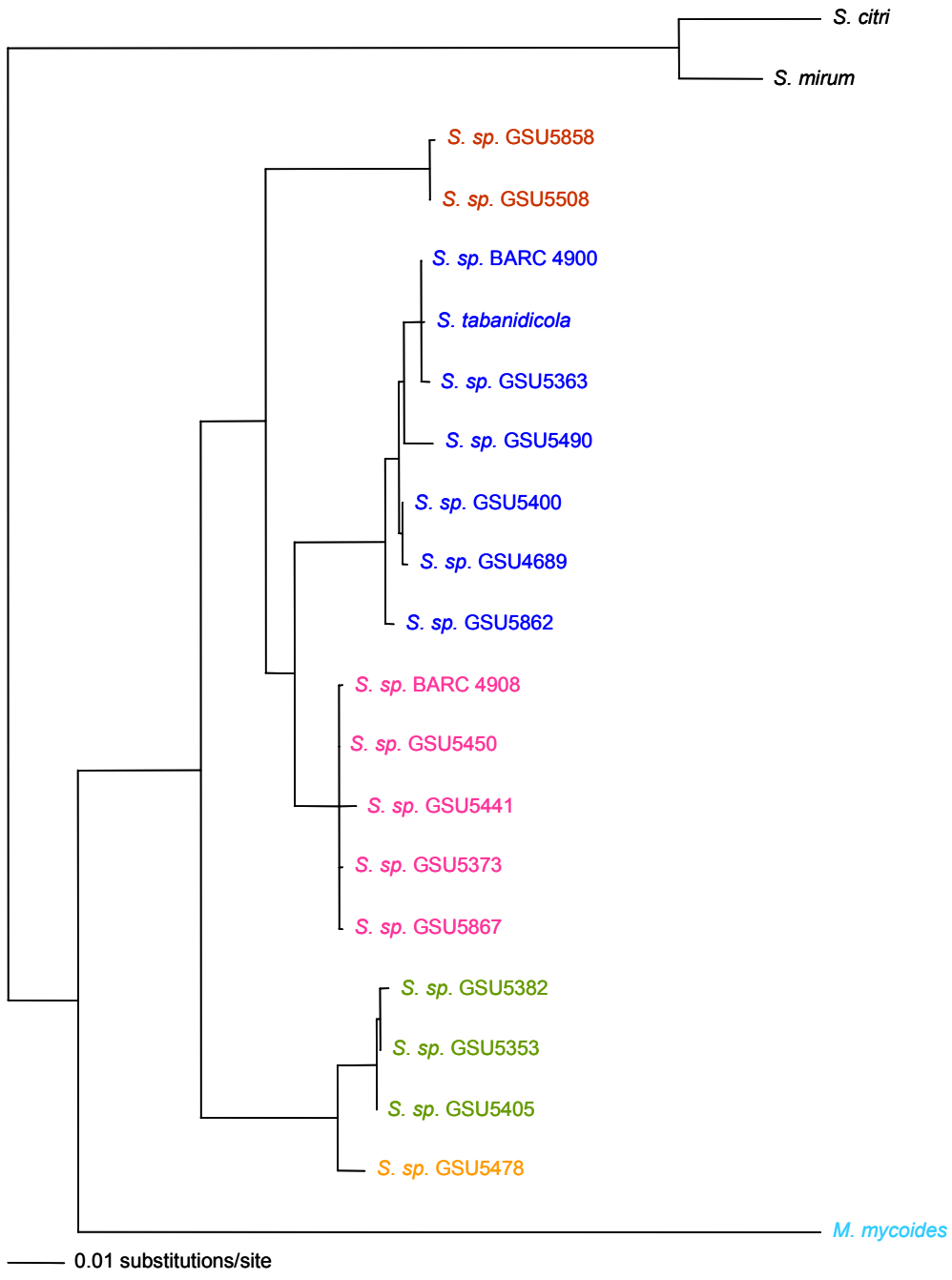

Figure 4C. Maximum Likelihood phylogenetic tree derived from partial 23S rDNA sequence analysis. The obtained 23S rDNA aligned matrix included a total of 1328 characters, *S. citri* and *S. mirum* were the out groups. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 5643.5080

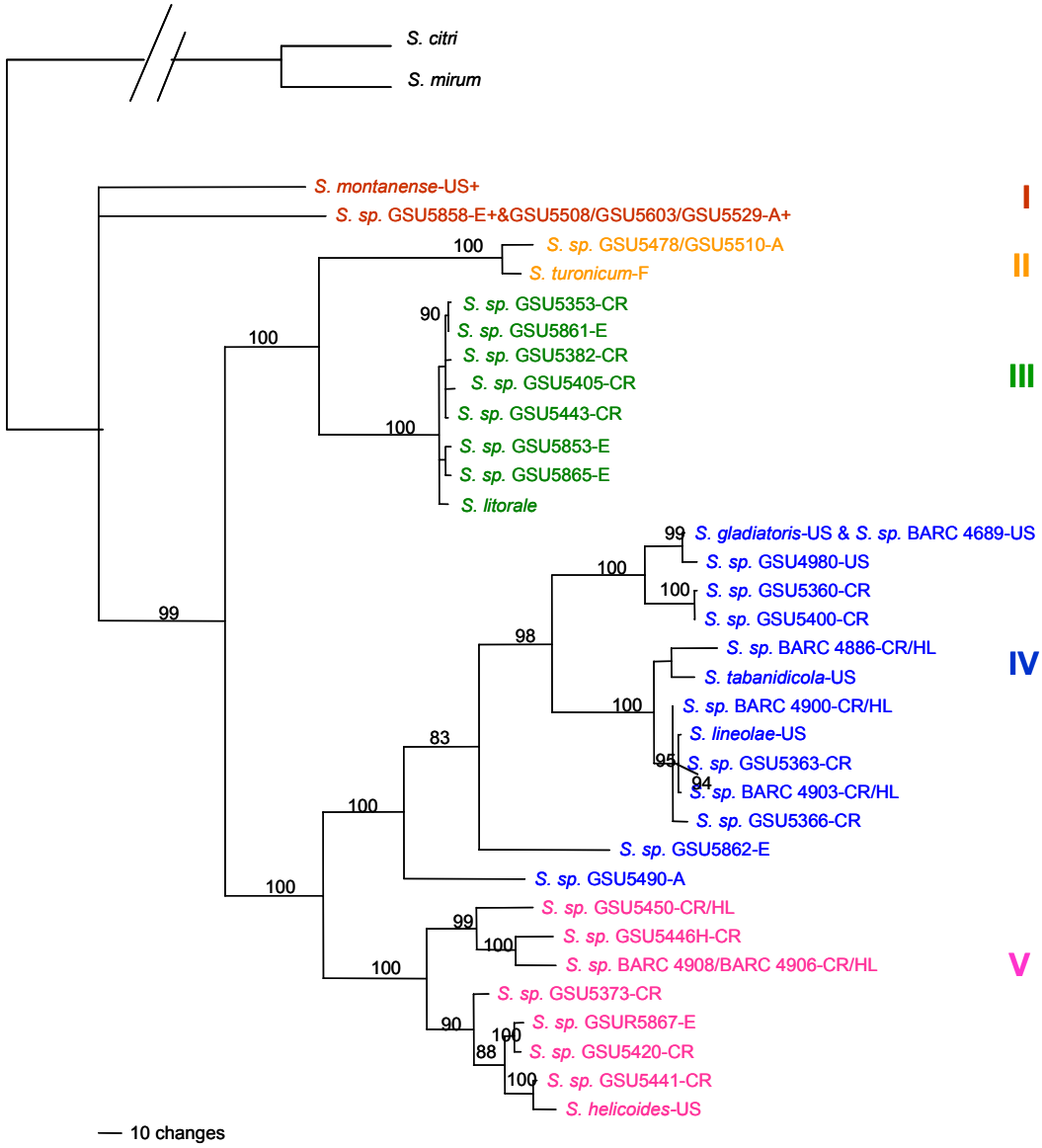

Figure 5A. Bayesian phylogenetic tree derived from partial *rpoB* sequence analysis. The obtained *rpoB* aligned matrix included a total of 1027 characters. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

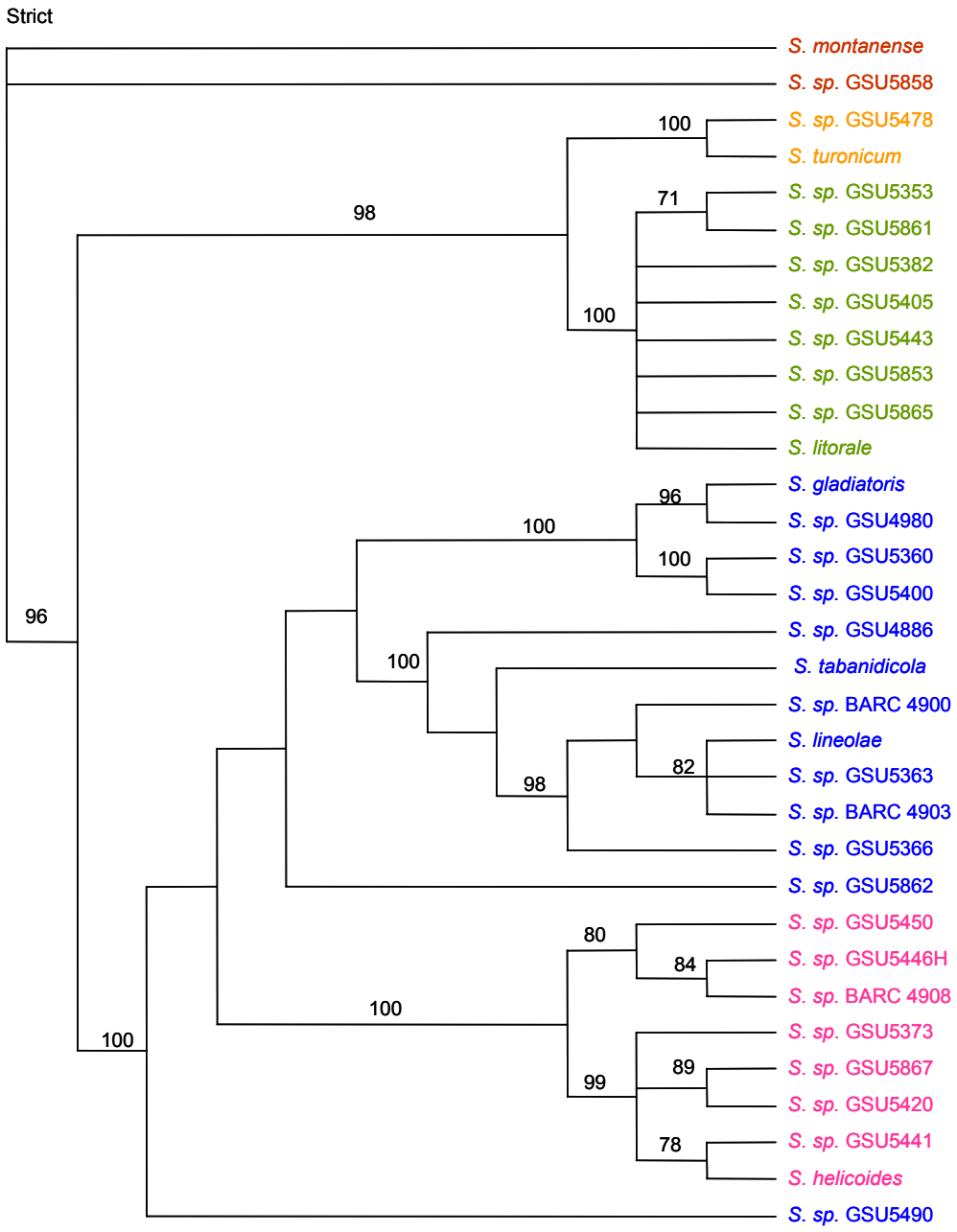

Figure 5B. Maximum Parsimony phylogenetic tree derived from partial *rpoB* sequence analysis. The obtained 16S rDNA aligned matrix included a total of 1027 characters. Maximum Parsimony heuristic search in PAUP found 32 best trees (length=1094, CI=0.550; RI=0.828; RC=0.456). 365 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

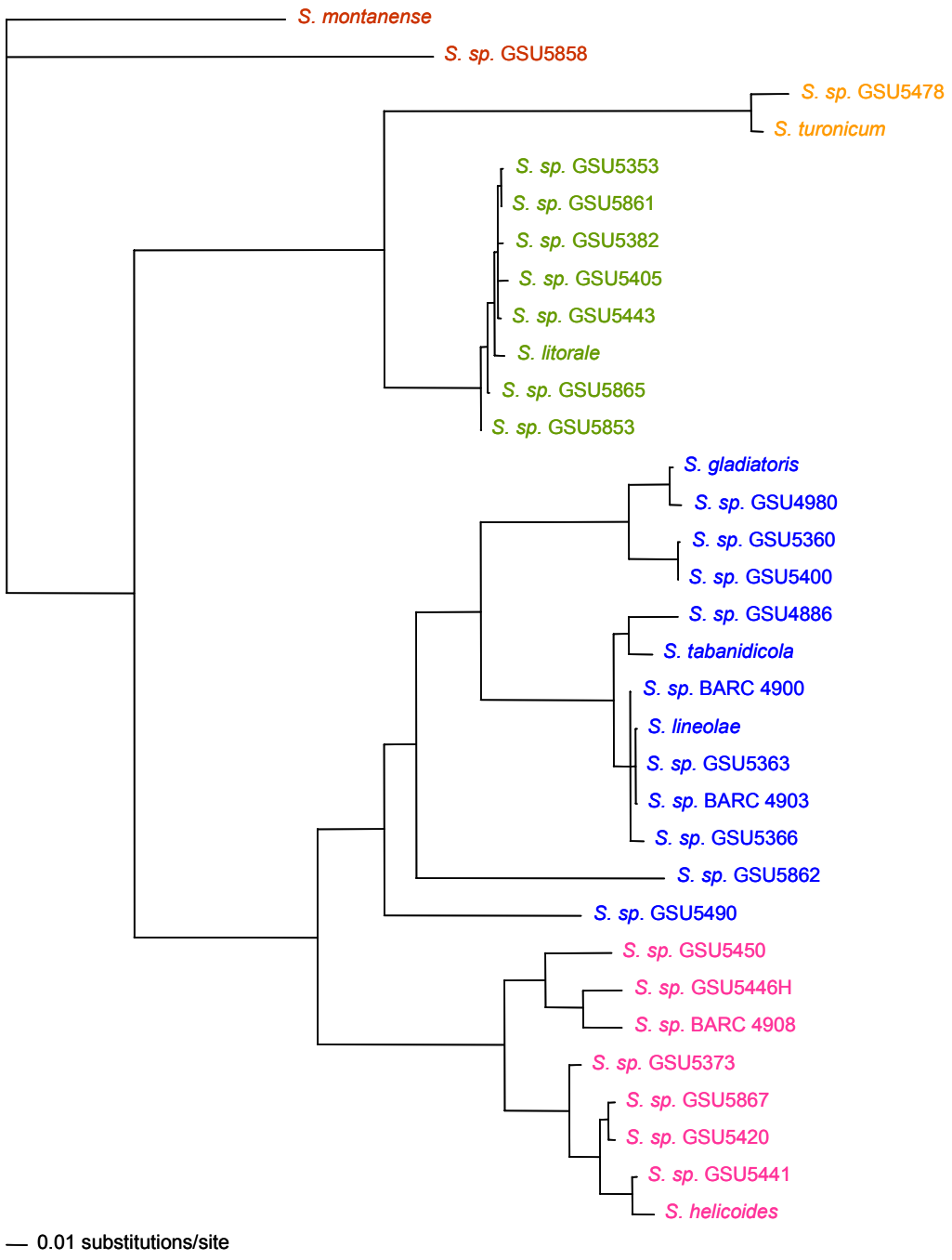

Figure 5C. Maximum Likelihood phylogenetic tree derived from partial *rpoB* sequence analysis. The obtained 16S rDNA aligned matrix included a total of 1027 characters. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 6406.54029.

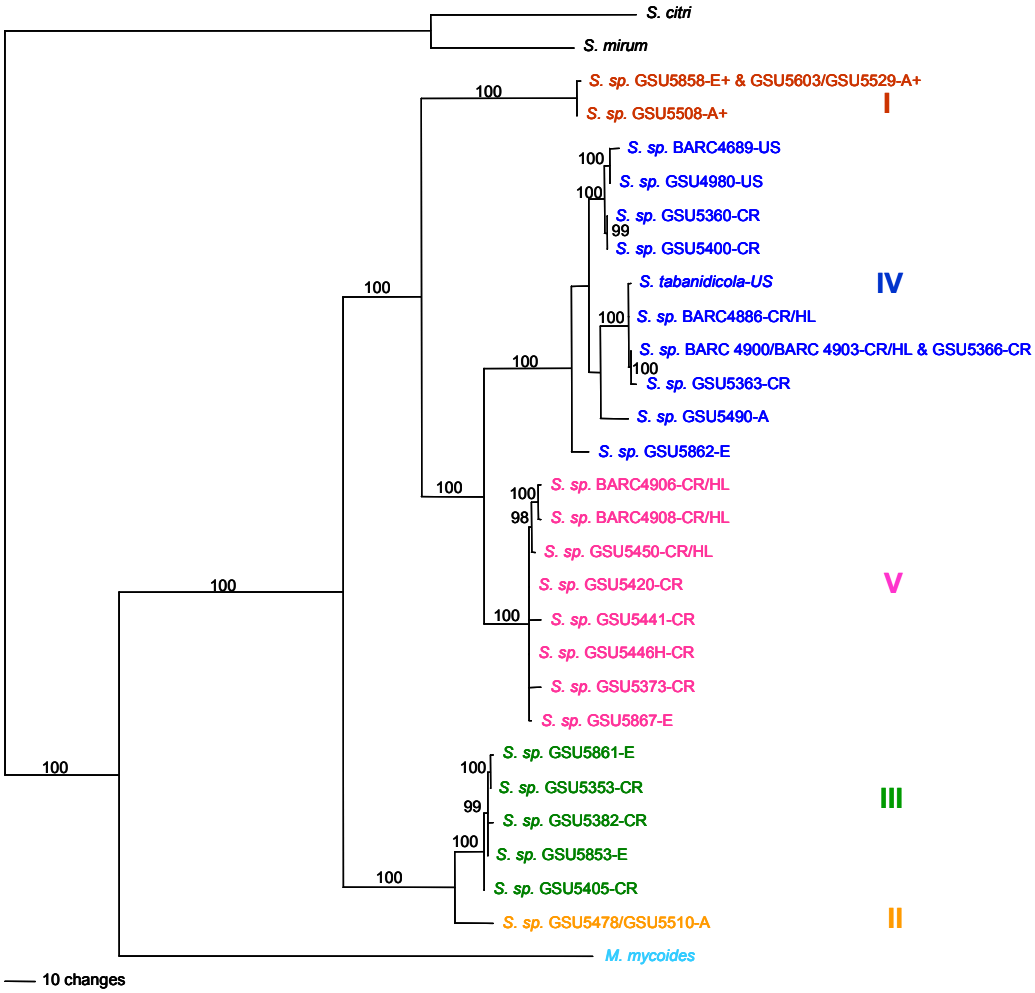

Figure 6A. Bayesian phylogenetic trees derived from concatenated 16S & 23S rRNA sequence analysis. The 16S rDNA gene sequences were concatenated to the 23S rRNA gene sequences for total evidence analysis of both datasets. The obtained concatenated 16S & 23S rRNA aligned matrix included a total of 2755 characters, *S. citri* and *S. mirum* were the out groups. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

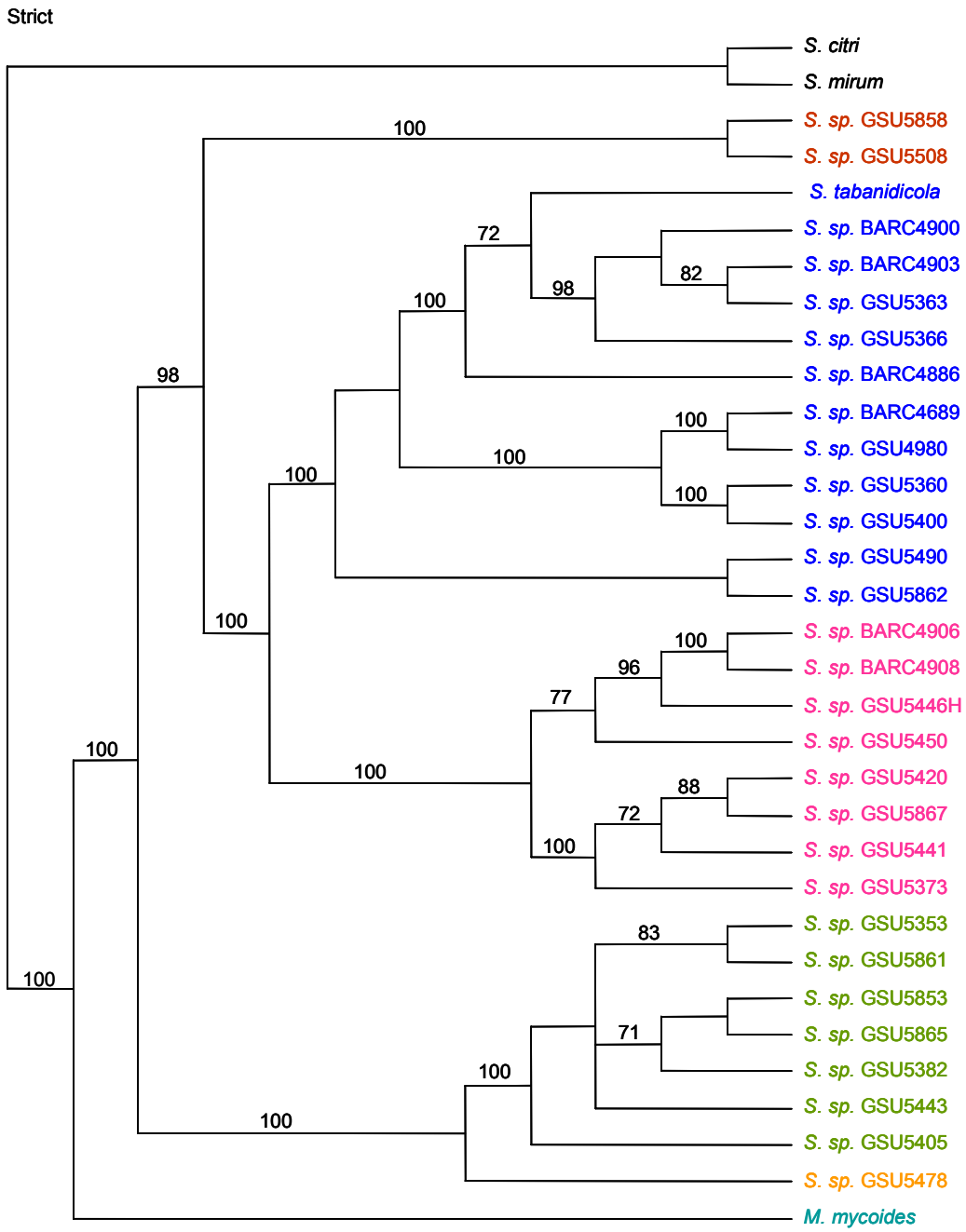

Figure 6B. Maximum Parsimony phylogenetic tree derived from concatenated 16S & 23S rRNA sequence analysis. The 16S rDNA gene sequences were concatenated to the 23S rRNA gene sequences for total evidence analysis of both datasets. The obtained concatenated 16S & 23S rRNA aligned matrix included a total of 2755 characters, *S. citri* and *S. mirum* were the out groups. Maximum parsimony heuristic search in PAUP found 8 best trees (length=806, CI=0.804; RI=0.885; RC=0.711). 361 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

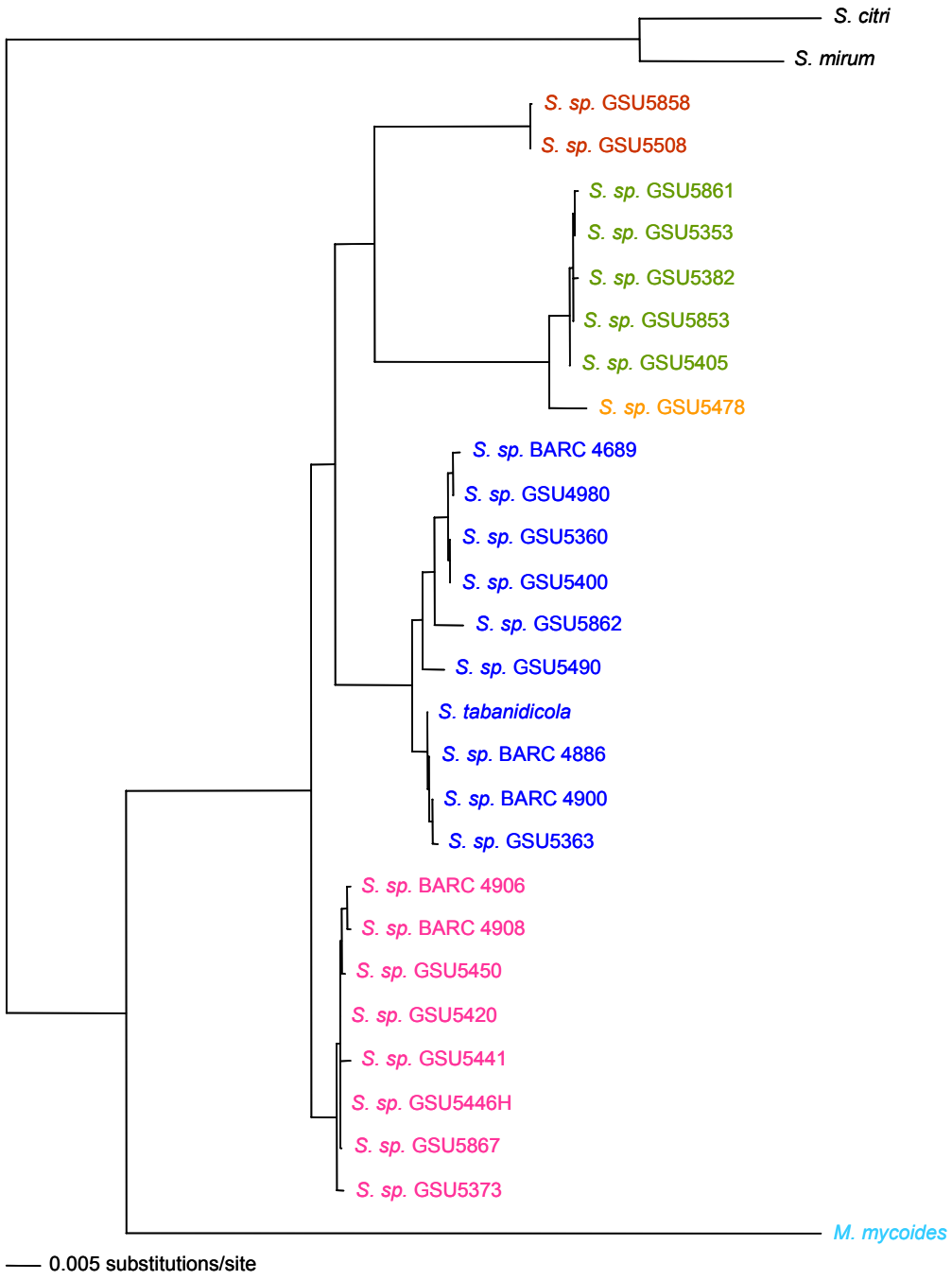

Figure 6C. Maximum Likehood phylogenetic tree derived from concatenated 16S & 23S rRNA sequence analysis. The 16S rDNA gene sequences were concatenated to the 23S rRNA gene sequences for total evidence analysis of both datasets. The obtained concatenated 16S & 23S rRNA aligned matrix included a total of 2755 characters, *S. citri* and *S. mirum* were the out groups. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 8065.01324.

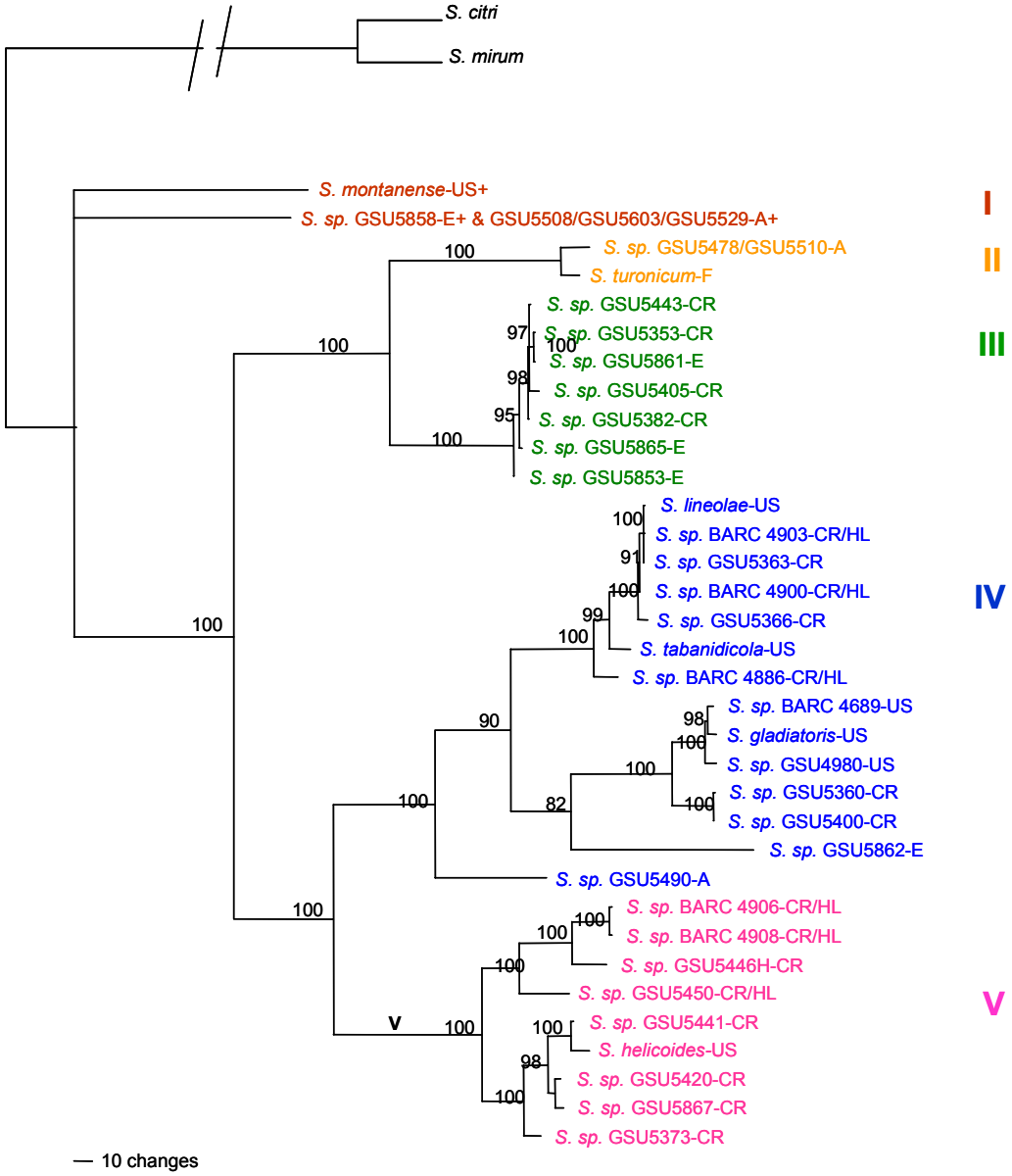

Figure 7A**.** Bayesian phylogenetic tree derived from concatenated 16S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S rRNA, ITS & *rpoB* sequence aligned matrix included a total of 2762 characters. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator.Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

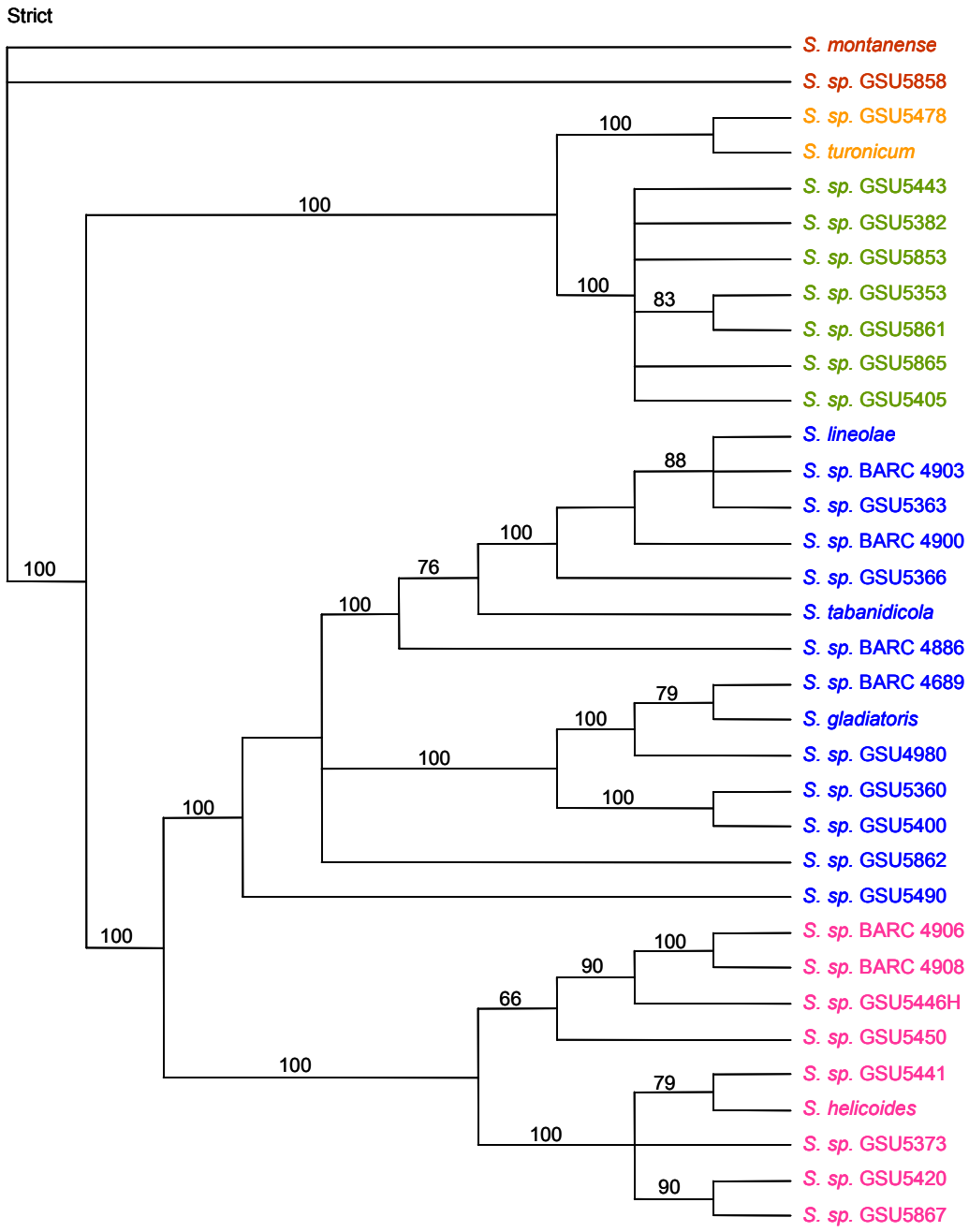

Figure 7B. Maximum Parsimony phylogenetic trees derived from concatenated 16S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S rRNA, ITS & *rpoB* aligned matrix included a total of 2762 characters. Maximum Parsimony heuristic search in PAUP found 16 best trees (length=1466, CI=0.583; RI=0.853; RC=0.497). 528 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

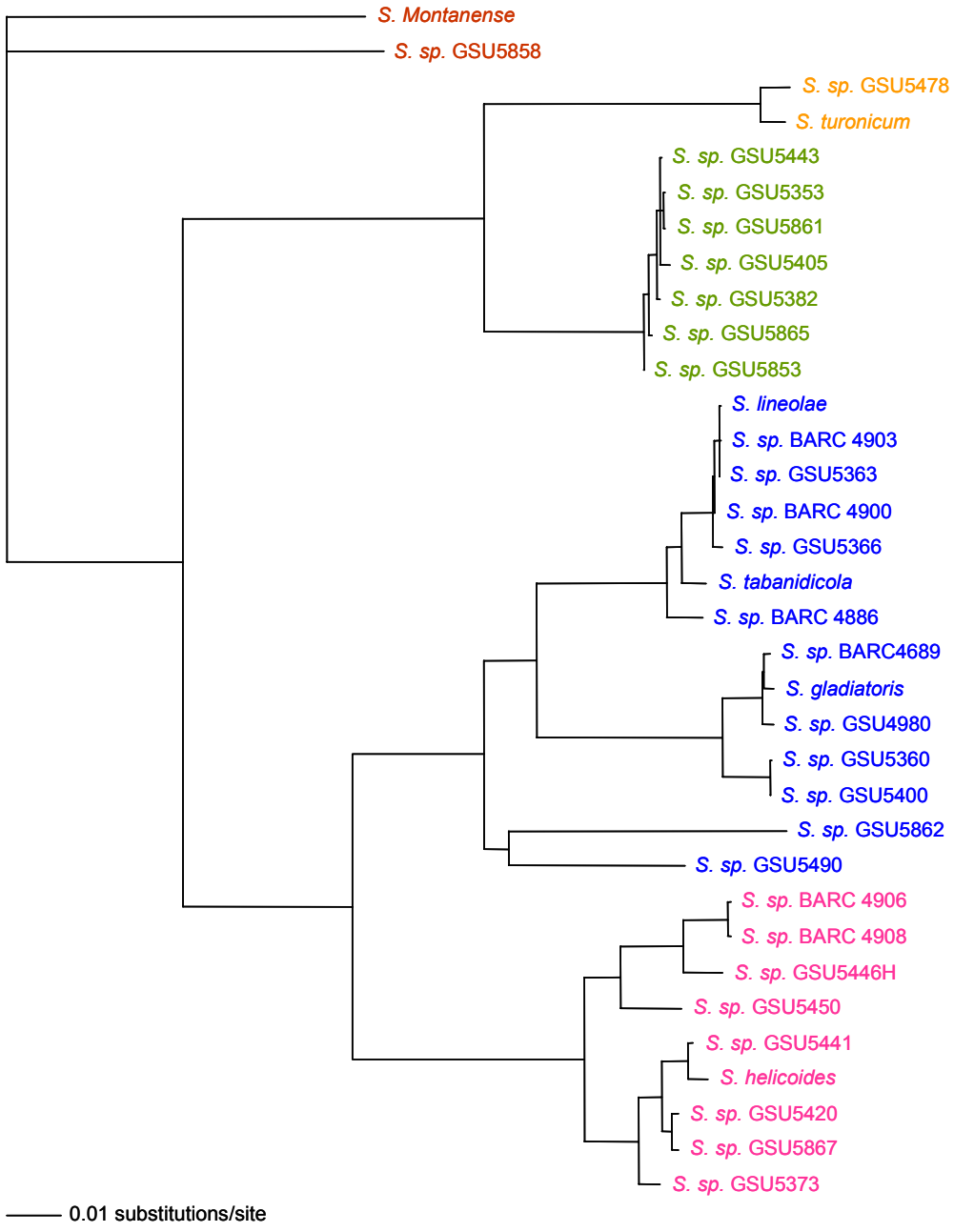

Figure 7C. Maximum Likelihood phylogenetic tree derived from concatenated 16S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S rRNA, ITS & *rpoB* aligned matrix included a total of 2762 characters. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 11079.00035

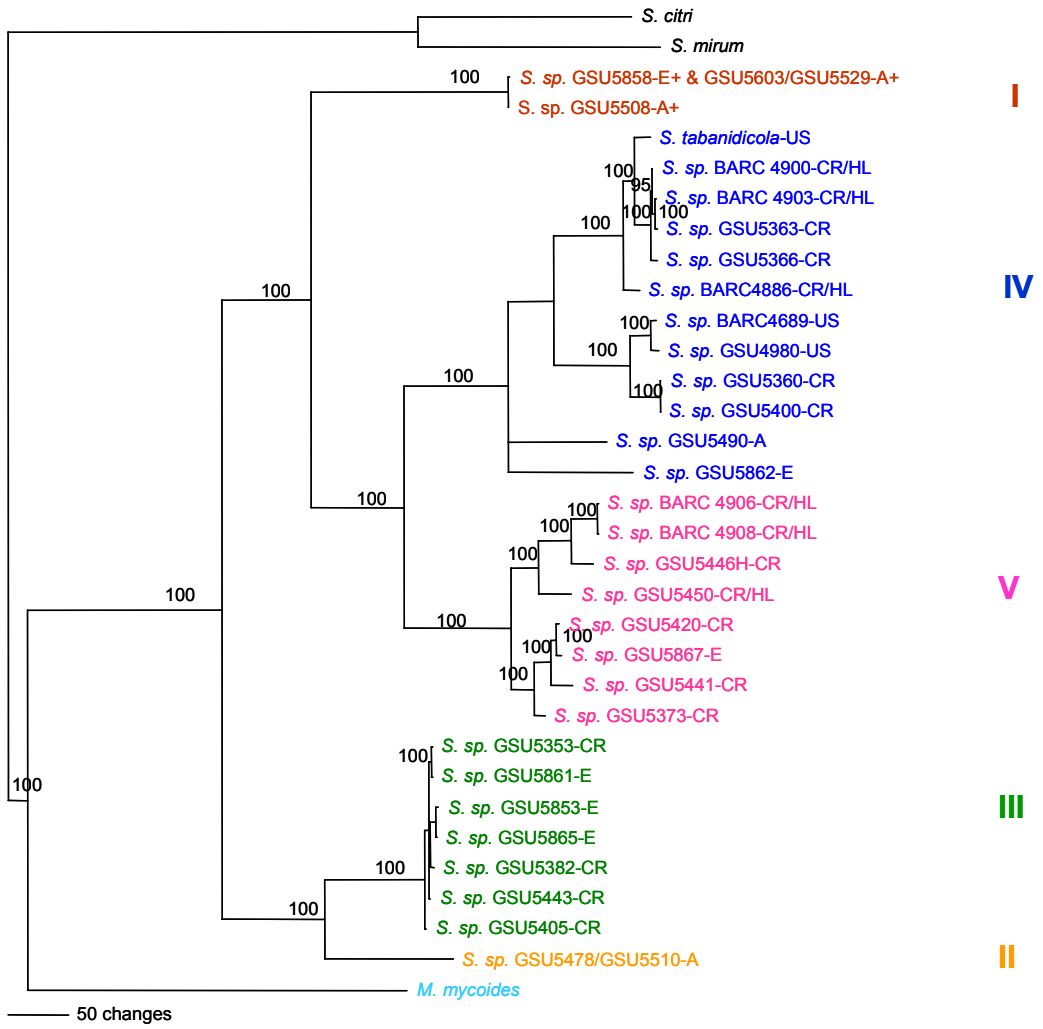

Figure 8A. Bayesian phylogenetic tree derived from concatenated 16S, 23S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S, 23S rRNA, ITS & *rpoB* aligned matrix included a total of 4089 characters. *S. citri* and *S. mirum* were the out groups. Isolate source information is as follows: A- Australia, E- Ecuador, CR- Costa Rica and CR/HL- Costa Rican High Lands, F- France, US- United States. The ability to hydrolyze arginine is indicated by a " + " after the locale indicator. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

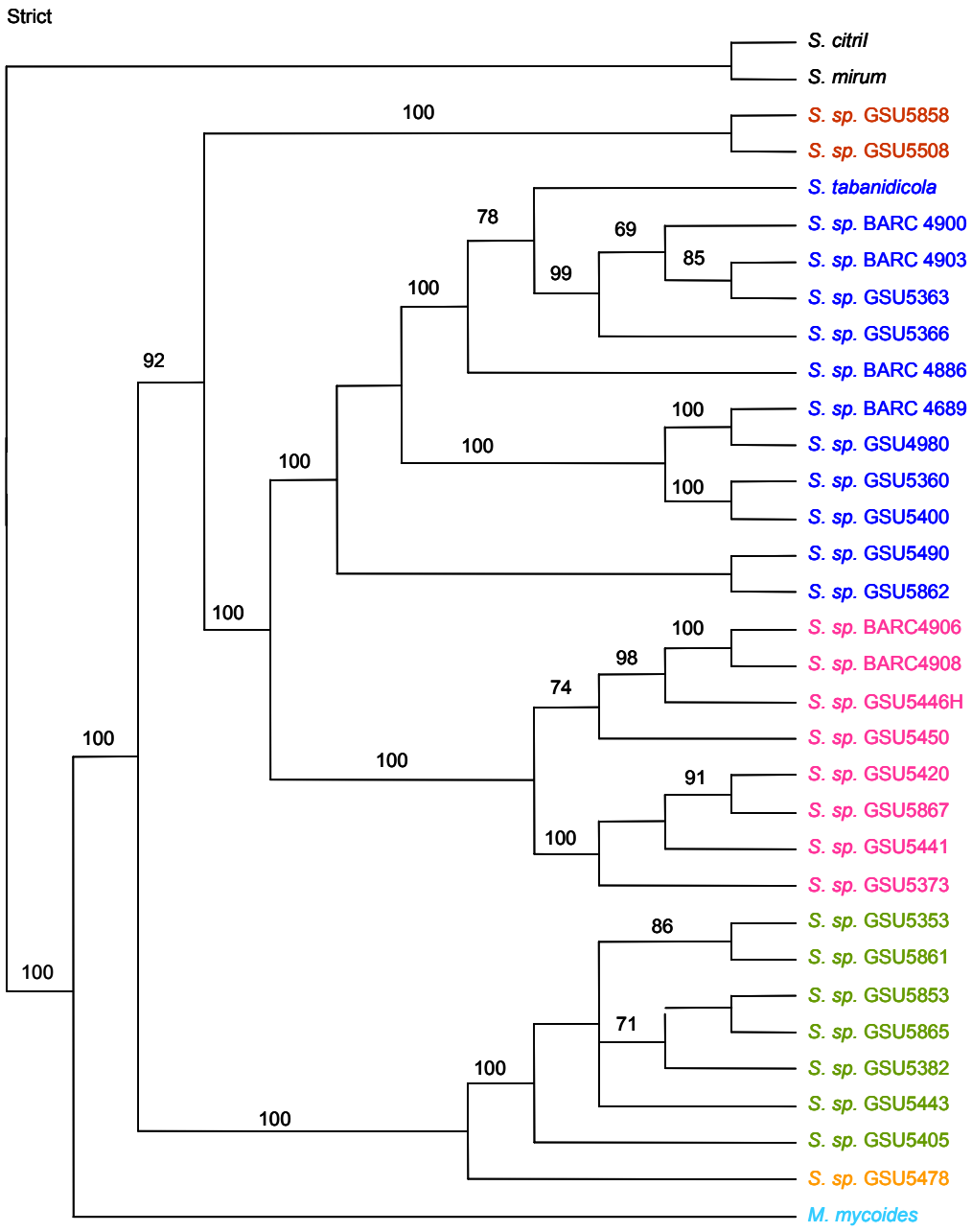

Figure 8B. Maximum Parsimony phylogenetic tree derived from concatenated 16S, 23S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S, 23S rRNA, ITS & *rpoB* aligned matrix included a total of 4089 characters. *S. citri* and *S. mirum* were the out groups. Maximum parsimony heuristic search in PAUP found 2 best trees (length=2661, CI=0.656; RI=0.829; RC=0.544). 909 characters are parsimony informative and branch support was evaluated by bootstrap analysis (100 replicates).

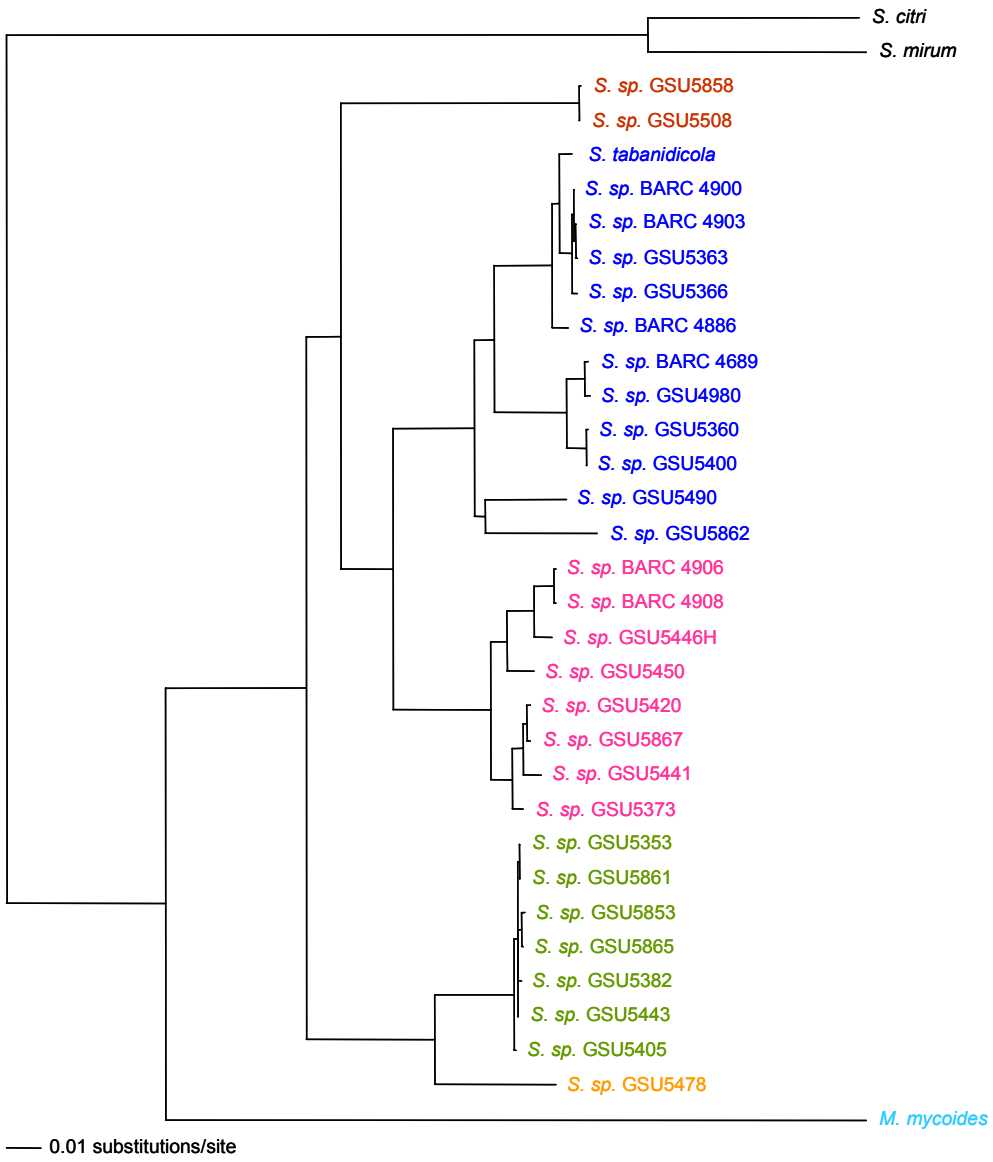

Figure 8C. Maximum Likelihood phylogenetic tree derived from concatenated 16S, 23S rRNA, ITS & *rpoB* sequence analysis. The obtained concatenated 16S, 23S rRNA, ITS & *rpoB* aligned matrix included a total of 4089 characters. *S. citri* and *S. mirum* were the out groups. Phylogram derived from a maximum likelihood analysis using the best model GTR+G+I and the –InL value is 18520.7990.

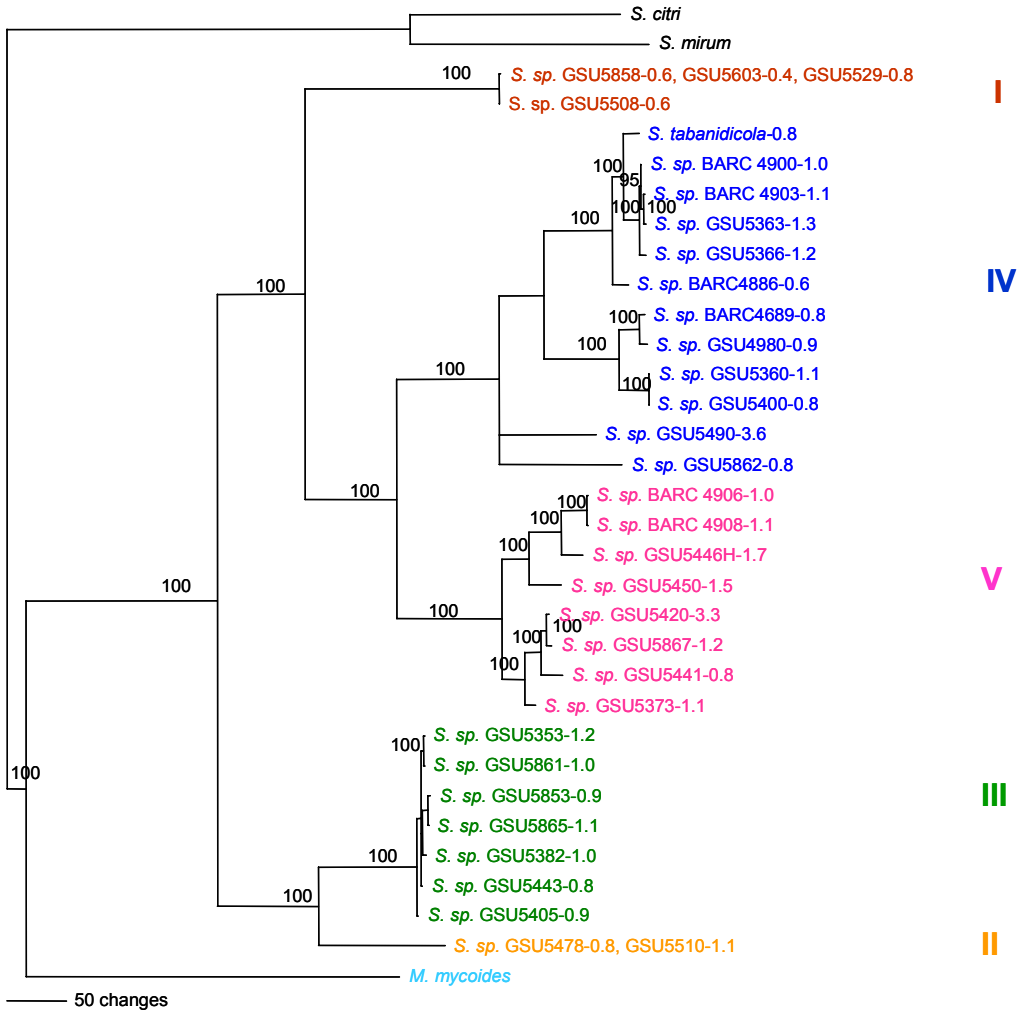

Figure 9. Optimal doubling time values overlayed on Bayesian phylogenetic tree derived from concatenated 16S, 23S rRNA, ITS & *rpoB* sequence analysis. The doubling time at the optimal growth temperature is shown for each novel isolate; growth profiles were determined as described by Konai et al. (1996). The obtained concatenated 16S, 23S rRNA, ITS & *rpoB* aligned matrix included a total of 4089 characters. *S. citri* and *S. mirum* were the out groups. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority rule consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

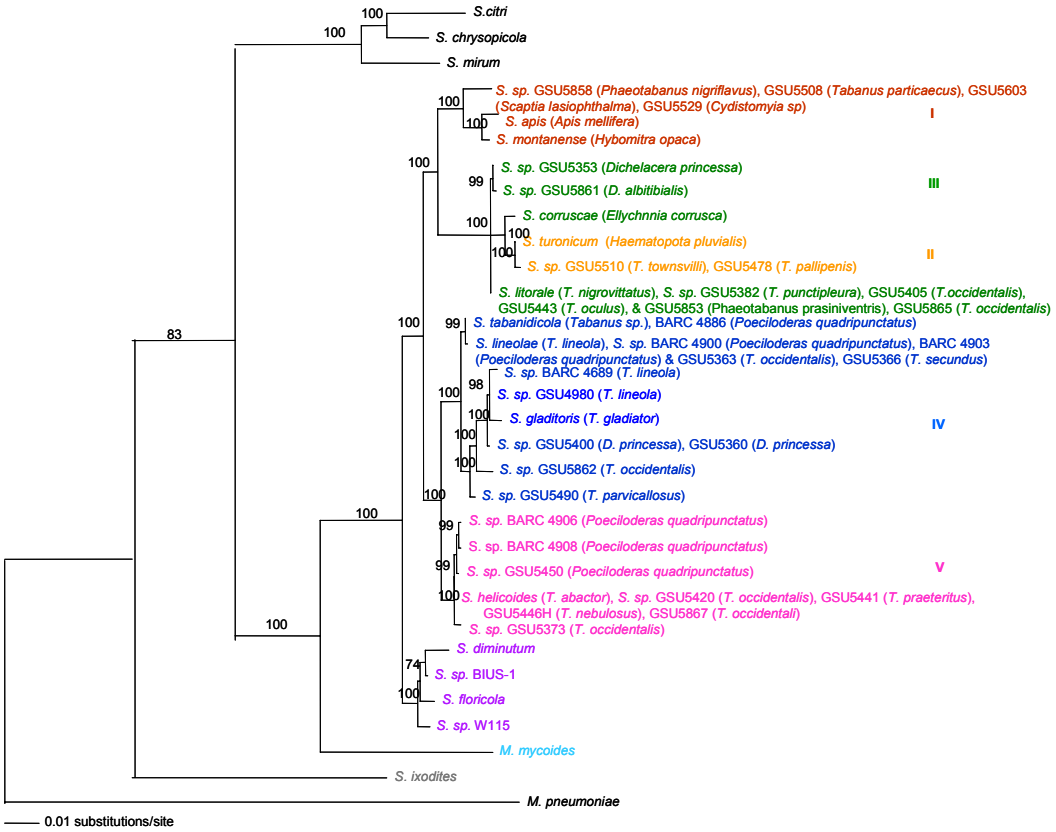

Figure 10. Spiroplasma host information overlayed on Bayesian phylogenetic tree derived from partial 16S rDNA sequence analysis. The host genus and species are indicated for each *Spiroplasma* isolate/species in the Apis Clade. The obtained 16S rDNA aligned matrix included a total of 1428 characters and *M. pneumoniae* was the out group. Bayesian analysis was performed in MrBayes by running 1,000,000 generations on 4 different chains. Posterior probability values shown were obtained by computing a 50% majority consensus of the trees generated by the search with the exclusion of those obtained before likelihood trees stabilized.

Table 6. Maximum parsimony heuristic search values for study genes.

Gene	Total characters	Best trees found	Informative characters	Length (score)	CI[1]	RI[2]	RC[3]
16S rRNA	1428	14	233	814	0.693	0.746	0.517
ITS	307	1	113	281	0.673	0.795	0.534
23S rRNA	1328	2	182	1214	0.911	0.807	0.736
rpoB	1027	32	365	1094	0.550	0.828	0.456
Conc[4]: 16S & 23S rRNA	2755	8	361	806	0.804	0.885	0.711
Conc: 16S rRNA, ITS & *rpoB*	2762	16	528	1466	0.583	0.853	0.497
Conc: 16S & 23S rRNA; ITS & *rpoB*	4089	2	909	2661	0.656	0.829	0.544

[1]CI= Consistency Index; [2]RI = Rescaled Index; [3]RC = Rescaled Consistency index
[4]Conc = Concatenated sequences for analysis

CHAPTER 4

DISCUSSION

The overall goal of this study was to examine the geographic diversification of spiroplasma isolates from three continents based on their genetic relatedness and to place these geographical isolates into a larger evolutionary context. Specifically, evolutionary relationships were resolved by generating 16S-23S rDNA spacer region, *rpoB* and 23S rRNA gene sequences for use in combination with the 16S rDNA sequence for each species. This study expanded the prior data set of eight tabanid-associated isolates from North America and Europe (Gasparich et al. 2004) to include new isolates from North America, Meso-America, South America and Australia.

Phylogenetic analyses of the novel isolates using maximum parsimony, maximum likelihood and bayesian analysis approaches generated trees that supported 5 major phylogenetic groups within the Apis Clade. Group I was defined by rapid growth and the ability to catabolyze arginine. Given the diverse hosts and host locales, biochemistry seemed to be the major determinant of group I. Groups II and III were closely related strain complexes that were basal with respect to the other groups. The individual strains grew well with average doubling times of 0.9-1.0 hr, and most isolates came from *Tabanus* sp. hosts. Group III contained North and South American strains. Surprisingly, group II had one French *Spiroplasma* species and 2 Australian isolates. Groups IV and V were also closely related strain complexes. Most isolates came from *Tabanus* sp. hosts in North and South America. The placement of GSU5490 from Australia into group IV was unexpected based on biogeography and growth rate (3.6 hr doubling time). On average,

group IV strains grew faster (average doubling time 1.1 hr) as compared to group V strains (average doubling time 1.5 hrs).

A total of 4 target sequences were used for phylogenetic analyses, 2 highly conserved (16S and 23S rRNA) and 2 more rapidly evolving (*rpoB* and ITS). The strongest deep branching support was generated by the 23S rRNA gene sequences, and the most defined resolution was obtained with the *rpoB* sequence. However, each individual gene/region contributed to the data analysis. The 16S rRNA trees aided in placement of the novel isolates into the Apis Clade and delineated their relationship to a large number of known species with available 16S rDNA sequences. The 23S rRNA gene, a highly conserved gene, resolved the deep branching position of groups II and III and also supported the five major Apis Clade groupings. The fast evolving ITS and *rpoB* sequences resolved the most recently diverging lineages. In particular, *rpoB* alone or in combination with other sequences helped to differentiate group III, IV and V. The results obtained with the ITS sequences were consistent with the other genes even though the parsimony informative charactes were less (113). The results from this study suggest that the use of all four genes significantly contributed to the resolution of the closely related spiroplasma isolates/species.

Non-genetic characters were used to enrich the phylogenetic analyses. Some characteristics were universal for all or nearly all of the spiroplasmas studied making them uninformative. The only informative biochemical property examined was the ability to catabolize arginine, as all study strains fermented glucose (LB. Regassa and D. Bostick, unpublished) and were resistant to penicillin (F.E. French, unpublished). The consistent optimal growth temperatures (30-32°C) were probably due to the common host

(Konai et al. 1996), so this parameter may be of utility for studies with more distantly related *Mollicutes*. In contrast, host, host locale, serology and growth rates varied for isolates and provided additional information. Overall host variability showed a lack of strict host specificity at the genus or species level. While some genuses were predominant (e.g. *Tabanus* sp.), they were not the sole host for phylogenetic or serological groups. The range of optimal growth rates (0.4-3.6 hrs) was small for the novel isolates, but phylogenetic patterns emerged with group I isolates growing the fastest (average doubling time 0.6 hr) and group V isolates the slowest (average doubling time 1.5 hr), (Konai et al. 1996). As discussed in more detail below, the serological data was congruent with the phylogenetic analyses and several large biogeographical patterns emerged.

Historically, surface serology has been used as a surrogate for DNA-DNA hybridization analyses when describing *Spiroplasma* species. Given the time consuming nature of these analyses, the field would like to move towards the use of genetic characters to delineate the species. Gasparich and co-workers (2004) generated 16S rDNA based trees, and demonstrated that evolutionary analyses did not contradict known serological relationships. Results from the current study also suggest that genetic characters exhibit congruency with serological data; however, the discriminatory power of surface serology is limited. The genetic characters are able to delineate more complex relationships that are not evident from the serological analyses. Complete analysis of the 8 serogroup pairs included in this study and analyses of the overall serological matrix will help determine if genetic characters from several genes/regions can reliably predict serogroups/species.

Isolates for this study were obtained from hosts in Australia, North America and South America. Initially, serological data suggested that spiroplasma species overlap for different regions; considerable serological crossover was observed between isolates from Costa Rica and the United States (Whitcomb et al. 2007). Similar overlap was seen in the phylogenetic analyses, with other general patterns also being observed. There was major overlap between North and South American isolates in groups III, IV and V; all Costa Rican highland isolates were confined to groups IV and V; and all but one Australian isolate (GSU5490) fell into groups I and II.

In summary, characterization of the novel isolates in this study added richness to the Apis Clade and enhanced our understanding of spiroplasma biodiversity and biogeography by examining evolutionary relationships and delineating geographical ranges.

REFERENCES

Bi K, Huang H, Gu W, Wang J, Wang W. 2008. Phylogenetic analysis of Spiroplasmas from three freshwater crustaceans (Eriocheir sinensis, Procambarus clarkia and Penaeus vannamei) in China. J Invertebr Pathol. Sep; 99(1):57-65.

Bové J M, C. Mouches, P. Carle-Junca, F. R. Degorce-Dumas, J. G. Tully & R. F. Whitcomb. 1983. Spiroplasmas of group I: The *Spiroplasma citri* cluster. *Yale J Biol Med* 56: 573-582.

Bove J M. 1997. Spiroplasmas: infectious agents of plants, arthropods and vertebrates Wien Klin Wochenschr. 109(14-15):604-12

Brown D. R., Whitcomb R.F and Bradbury J M. 2007. Revised minimal standards for description of new species of the class *Mollicutes* (division *Tenericutes*) Int J Syst Evol Microbiol **57**; 2703-2719.

Calavan E C & J. M. Bové. 1989. Ecology of *Spiroplasma citri.* In: The mycoplasmas. Eds: Whitcomb RF, Tully JG, Academic Press, San Diego, CA.

Carle, P., Laigret, F., Tully, J.G., and Bové, J.M. 1995. Heterogeneity of genome sizes within the genus *Spiroplasma*. Int. J. Syst. Bacteriol. 45:178-181.

Clark T B. 1978. Honey bee spiroplasmosis a new problem for bee keepers. Amer Bee J118, 18-23.

Clark T B, R. B. Henegar, L. Rosen, K. J. Hackett, R. F. Whitcomb, J.E. Lowry, C. Saillard, M. Bové, J.G. Tully & D.L. Williamson. 1987. New spiroplasmas from insects and flowers: isolation, ecology, and host association. *Isr J Med Sci* 23, 687-690

Clark T B. and R. F. Whitcomb. 1984. Pathogenicity of mollicutes for insects: possible use in biological control. Ann Microbiol (Paris).135A (1):141-50.

Davis R E 1978. Spiroplasma associated with flowers of the tulip tree (*Liriodendron tulipifera L.*). *Can J Microbiol* 24, 954-959.

Ebbert M A & L. R. Nault. 1994. Improved overwintering ability in *Dalbulus maidis* (Homoptera: *Cicadellidae*) vectors infected with *Sprioplasma kunkelii* (*Mycoplasmatales: Spiroplasmataceae*). *Environ Ent* 23, 634-644.

Fukatsu and Nikoh. 1998. Two Intracellular Symbiotic Bacteria from the Mulberry Psyllid *Anomoneura mori* (Insecta, Homoptera) Applied and Environmental Microbiology, p. 3599–3606

Gasparich G E, C. Saillard, E. A. Clark, M. Konai, F. E. French, J. G. Tully, K. J. Hackett & R. F. Whitcomb. 1993. Serologic and genomic relatedness of group VIII and group XVII spiroplasmas and subdivision of spiroplasma group VIII into subgroups. *Int J Syst Bacteriol* 43, 338-341.

Gasparich, G.E., Whitcomb, R.F., Dodge, D., French, F.E., Glass, J., and Williamson, D.L. 2004. The genus Spiroplasma and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma mycoides clade. Int J Syst Evol Microbiol, 54: 893-918.

Hackett, K.J., and Clark, T.B. 1989. Spiroplasma Ecology. *In* The mycoplasmas. Vol. 5. *Edited by* R.F. Whitcomb and J.G. Tully. Academic Press, San Diego, Calif. pp. 113-20.

Hackett, K.J., Whitcomb, R.F., Tully, J.G., Lloyd, J.E., Anderson, J.J., Clark, T.B., Henegar, R.B., Rose, D.L., Clark, E.A., and Vaughn, J.L. 1992. *Lampyridae* (*Coleoptera*): A Plethora of mollicute associations. Microbial. Ecol. 23: 181-193.

Hackett K J, R. F. Whitcomb, F. E. French J. G. Tully, G. E. Gasparich, D. L. Rose, P. Carle, J. M. Bové, R. B. Henegar, T. B. Clark, M. Konai, E. A. Clark & D. L. Williamson. 1996a. *Spiroplasma corruscae* sp. nov., from a firefly beetle (Coleoptera: Lampyridae) and tabanid flies (Diptera: Tabanidae). *Int J Syst Bacteriol* 46, 947-950.

Hackett K J, R. F. Whitcomb, T. B. Clark, R. B. Henegar, D. E. Lynn, A. G. Wagner, J. G. Tully, G. E. Gasparich, D. L. Rose, P. Carle, J. M. Bové, M. Konai, E. A. Clark, J. R. Adams & D. L. Williams. 1996. *Spiroplasma leptinotarsae* sp. nov., from the Colarado potato beetle *Leptinotarsa decemlineata. Int J Syst Bacteriol* 46, 906-911.

Hélias, C., Vazeille-Falcoz, M., Le Goff, F., Abalain-Colloc, M.-L., Rodhain, F., Carle, P., Whitcomb, R.F., Williamson, D.L., Tully, J.G., Bové, J.M. and Chastel, C. 1998. *Spiroplasma* turonicum sp. nov. from Haematopota horse flies (Diptera, Tabanidae) in France. Int. J. Syst. Bacteriol. 48,457-461.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

International Committee on Systematic Bacteriology. Subcommittee on the Taxonomy of Mollicutes. 1995. Revised minimal standards for descriptions of new species of the class Mollicutes (Division Tenericutes). Int J Syst Bacteriol 45, 605-612.

Johansson, K.-E. and Pettersson, B. 2002. Taxonomy of *Mollicutes*. *In* Molecular biology and pathogenicity of *Mycoplasmas*. *Edited by* S. Razin and R. Herrmann. Kluwer Academic/Plenum: London, UK, pp. 1-29.

Junca P, C. Saillard, J. G. Tully, O. Garcia-Jurado, J-R. Degorce-Dumas, C. Mouches, J-C. Vignault, R. Vogel, R. McCoy, R. F. Whitcomb, D. L. Williamson, J. Latrille & J. M. Bové. 1980. Caractérisation de spiroplasmes isolés d'insectes et de fleurs de France continentale, de Corse et du Maroc. Proposition pour une classification des spiroplasmes. *C R Acad Sci Paris Ser D* 290, 1209-1211.

Kirchhoff H, Heitmann J, Trautwein G. Pathogenicity of Spiroplasma sp. strain SMCA in rabbits. 1981. clinical, microbiological, and histological aspects. Infect Immun. 33(1):292-6.

Konai M, E. A. Clark, M. Camp, A. L. Koch & R. F. Whitcomb. 1996. Temperature ranges, growth optima and growth rates of *Spiroplasma* (*Spiroplasmataceae*, class *Mollicutes*) species. *Curr Microbiol* 32, 314-319.

Lorenz B, Schroeder J, Reischl U. 2002. First evidence of an endogenous Spiroplasma sp. infection in humans manifesting as unilateral cataract associated with anterior uveitis in a premature baby. Graefes Arch Clin Exp Ophthalmol. May;240(5):348-53.

Maddison, W. P. & Maddison, D. R. 1992. MacCLade: analysis of phylogeny and character evolution, version 3.0. Sunderland, MA: Sinauer Associates.

McCoy R E, D. S. Williams & D. L. Thomas. 1979. Isolation of mycoplasmas from flowers. In Proceedings of the US-ROC plant mycoplasma seminar. National Science Council, Taipei, Taiwan 75-81.

Megraud F, Gamon LB, McGarrity GJ. 1983. Characterization of Spiroplasma mirum (suckling mouse cataract agent) in a rabbit lens cell culture. Infect Immun. 42(3):1168-75.

Mouches C, A. Menara, J. G. Tully, D. L. Rose, R. E. McCoy, P. Clarle-Junca, M. Granier & C. Saillard. 1983. *Spiroplasma apis*, a new species from the honeybee (*Apis mellifera*). *Ann Inst Microbiol (Paris)* 134A, 383-397.

Moulder RW, French FE, Chang CJ. 2002. Simplified media for spiroplasmas associated with tabanid flies Can J Microbiol. 48(1):1-6

Nunan L M, C. R. Pantoja, M. Salazar, F. Aranguren & D. V. Lightner. 2004. Characterization and molecular methods for detection of a novel spiroplasma pathogenic to *Penaeus vannamei*. *Dis Aquat Organ* 62, 255-64.

Nunan L M, D. V. Lightner, M. Oduori & G. E. Gasparich. 2005. *Spiroplasma penaei* sp. nov. associated with mortalities in *Penaeus vannamei*, Pacific white shrimp. *J Syst Evol Microbiol* Papers in Press, published online 1 July.

Posada D and Crandall KA 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9): 817-818.

Regassa, L.B., K. Stewart, A. Murphy, F. French, T. Lin and R.F. Whitcomb. 2004. Digging out of a Taxonomic Quandary: Differentiation of Group VIII *Spiroplasma* Strains with Sequences of the 16S-23S rDNA Intergenic Spacer Region. Can. J. Microbiol. 50:1061-1067.

Regassa, L.B. and Gasparich, G.E. 2006. Spiroplasmas: evolutionary relationships and biodiversity. Frontiers in Bioscience 11:2983-3002.

Rogers, M.J., Simmons, J., Walker, R.T., Weisburg, W.G., Woese, C.R., Tanner, R.S Robinson, I.M., Stahl, D.A., Olsen, G., Leach, R.H. and Maniloff, J. 1985. Construction of the mycoplasma evolutionary tree from 5S rRNA sequence data.Proc. Natl. Acad. Sci. USA 82: 1160-1164.

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rose D L, J. G. Tully, J M. Bové & R. F. Whitcomb. 1993. A test for measuring growth responses of mollicutes to serum and polyoxyethylene sorbitan. *Int J Syst Bacteriol* 43, 527-532.

Saglio P, M. L' Hospital, D. Lafliche, G. Dupont, J. M. Bové, J. G. Tully & E. A. Freudt. 1973. *Spiroplasma citri* gen. and sp. N.: A mycoplasma-like organism associated with "stubborn" disease of citrus. *Int J Syst Biol* 23, 191-204.

Swofford, D.L. 2001. PAUP: phylogenetic analysis using parsimony (and other methods). Ver. 4. Sinauer Associates, Sunderland, Massachusetts.

Trachtenberg S. 2004. Shaping and moving a *Spiroplasma*. *J Mol Microbiol Biotechnol* 7, 78-87.

Tully J G, R. F. Whitcomb, D. L. Rose & J. M. Bové. 1982. *Spiroplasma mirum*, a new species from the rabbit tick (*Haemaphysalis leporispalustris*). *Int J Syst Bacteriol* 32, 92-100.

Tully J G. 1983. Cloning and filtration techniques for mycoplasmas. *Methods Mycoplasmol*. 1, 173-177.

Tully J G, D. L. Rose, E. Clark, P. Carle, J. M. Bové, R. B. Henegar, R. F. Whitcomb, D. E. Colflesh & D. L. Williamson. 1987. Revised group classification of the genus *Spiroplasma* (class *Mollicutes*), with proposed new groups XII to XXIII. *Int J Syst Bacteriol* 37, 357-364.

Tully J G & R. F. Whitcomb. 1991. The genus *Spiroplasma*. In: The prokaryotes. Eds: Starr M P, Stolp H, Truper H G, Balows A, Schlegel H G, Springer, NY.

Tully J G, D. L. Rose, C. E. Yunker, P. Carle, J. M. Bové, D. L. Williamson & R. F. Whitcomb. 1995. *Spiroplasma ixodetis* sp. nov., a new species from *Ixodes pacificus* ticks collected in Oregon. *Int J Syst Bacteriol* 45, 23-28.

Volokhov DV, Neverov AA, George J, Kong H, Liu SX, Anderson C, Davidson MK, and Chizhikov V. 2007. Genetic analysis of housekeeping genes of members of the genus Acholeplasma: phylogeny and complementary molecular markers to the 16SrRNA gene. Molecular phylogenetics and evolution; 44 (2):699-710

Wang W, B. Wen, G. E. Gasparich, N. Zhu, L. Rong, J. Chen & Z. Xu. 2004. A spiroplasma associated with tremor disease in the Chinese mitten crab (*Eriocheir sinensis*). *Microbiol* 150, 3035-3040.

Wang W, W. Gu, Z. Ding, Y. Ren, J. Chen & Y. Hou. 2005. A novel *Spiroplasma* pathogen causing systemic infection in the crayfish *Procambarus clarkii* (Crustacea: Decapod), in China. *FEMS Microbiol Lett* 249, 131-137.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., Van Etten, J., Maniloff, J., and Woese, C.R. 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. 171: 6455-6467.

Whitcomb R F & J. G. Tully. 1984. Family *Spiroplasmastaceae*, genus *Spiroplasma*. In: Bergey's manual of systematic bacteriology, vol. 1, Ninth Edition. Eds: Krieg N R, Holt J G, Williams D L, Wilkins M D, Williams and Wilkens, Baltimore .

Whitcomb R F, T. A. Chan, D. L. Williamson, C. Liao, J. G. Tully, J. M. Bové, C. Mouches, D. L. Rose, M. E. Coan & T. B. Clark. 1986. *Spiroplasma kunkelli* sp. Nov.: Characterization of the etiological agent of corn stunt disease. *Int J Syst Bacteriol* 36, 170-178.

Whitcomb R F & K. J. Hackett. 1987. Cloning by limiting dilution in liquid media: an improved alternative for cloning mollicute species. *Isr J Sci* 23, 517.

Whitcomb R F, F. E. French, J. G. Tully, G. E. Gasparich, D. L. Rose, P. Carle, J. M. Bové, R. B. Henegar, M. Konai, K. J. Hackett, J. R. Adams, T. B. Clark & D. L. Williamson. 1997. *Spiroplasma chrysopicola* sp. nov., *Spiroplasma gladiatoris* sp. nov., *Spiroplasma helicoides* sp. nov., and *Spiroplasma tabanidicola* sp. nov. from tabanid (Diptera: Tabanidae) flies. *Int J Syst Bacteriol* 47, 713-719.

Whitcomb, R.F.; Tully, J.G.; Gasparich, G.E.; Regassa, L.B.; Williamson, D.L.; French, F.E. 2007. *Spiroplasma* species in the Costa Rican highlands: implications for biogeography and biodiversity. Biodiversity and Conservation 16: 3877-3894.

Williamson D L, R. F. Whitcomb & J. G. Tully. 1978. The Spiroplasma deformation test, a new serological method. *Curr Microbiol* 1, 203-207.

Williamson D L, J. G. Tully & R. F. Whitcomb. 1979. Serological relationships of spiroplasmas as shown by combined deformation and metabolism inhibition tests. *Int J Syst Bacteriol* 29, 345-351.

Williamson D L, J. G. Tully & R. F. Whitcomb. 1989. The genus *Spiroplasma.* In: The mycoplasmas. Eds: Whitcomb RF & Tully JG. Academic Press, CA.

Williamson D L, R. F. Whitcomb, J. G. Tully, G. E. Gasparich, D. L. Rose, P. Carle, J. M. Bové, K. J. Hackett, J. R. Adams, R. B. Henegar, M. Konai, C. Chastel & F. E. French. 1998. Revised group classification of the genus *Spiroplasma*. *Int J Bacteriol* 48, 1-12

Williamson, D.L., Adams, J.R., Whitcomb, R.F., Tully, J.G. Carle, P., Konai, M., Bové, J.M., and Henegar, R.B. 1997. *Spiroplasma platyhelix* sp. nov., a new mollicute with unusual morphology and genome size from the dragonfly *Pachydiplax longipennis*. Int J Syst Bacteriol 47, 763-766

Williamson D L, B. Sakaguchi, K. J. Hackett, R. F. Whitcomb, J. G. Tuly, P. Carle, J. M. Bové, J. R. Adams, M. Konai & R. B. Henegar. 1999. *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *Int J Syst Bacteriol* 49, 611-618.

Woese, C.R., Maniloff, J., and Zablen, L.B. 1980. Phylogenetic analysis of the mycoplasmas. PNAS USA 77: 494-498.

Appendix A: Partial characterization of strain GSU5358 from Australia.

Gene	Sequence
16S rRNA	TCCTACCTTGGTACGCTCTCTCCTTACGGTTAAGATACGTGCTTCTGGTATTACCAACTCTCATGGTGT GACGGGCGGTGTGTACAAGACCCGAGAACGTATTCACCGCGACGTTGCTGATTCGCGATTACTAGTGA TTCCGGCTTCATGAAGTCGAGTTGCAGACTTCAATCCGAACTGAGACTGACTTTTTGAGATTAGCTCCC CCTCACAGGATTGCGACTCTTTGTATCAGCCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCAT GATGATTTGACGTCATCCCCACCTTCCTCTAGCTTACACTAGCAGTCTCGTTAAAGTGCTCAACTTAAT GTTAGTAACTAACGATAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACACCTCACGGCACGAGCTGA CGACAACCATGCACCACCTGTCTAATGTTGGCCTCCACTACATCTCTGTAGTTTTGCACTGGATGTCAA GCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCACTTGTGCGGGTCCCCGTC AATTCCTTTGAGTTTCACTCTTGCGAGCATACTACTCAGGCGGAGTACTTAATGCGTTAGCTGCAGCAC CGACAAATTGCCGACACTTAGTACTCAACGTTTACGGCGTGGACTACTAGGGTATCTAATCCTATTTGC TCCCCACGCTTTCGTGCCTCAGTGTCAATCACAGGCCAGTAGACCGCCTTCGCCACTGGTGTTCCTCCA TATATCTACGCATTCCACCGCTACACATGGAATTCCATCTACCTCTCCTGTATTCTAGTTAGCCAGTTTT CAAGGCGGACCAGGGTTGAGCCCTGGGATTTAACCTCAAACTTAACTAACCACCTACGCACCCTATAC GCCCAATAAATCCGGATAACGCTTGCCACCTATGTATTACCGCGGCTGCTGGCACATAGTTAGCCGTG GCTTTCTGGTAAGGTACCGTCAAACTAAAAGCATTTCCTCTTCTAGCATTTCTTCCCTTACAACAGAGC TTTACAATCCGAAGACCTTCATCACTCACGCGGCATTGCTTCATCAGGCTTTCGCCCATGTGAAAAATT CCCTACTGCTGCCTCCCGTAGGAGTCTGGGCCGTATCTCAGTCCCAATGTGGCCGATCAACCTCTCAGT TCGGCTACGTATCATCGCCTAGGTGAGCCATTACCTCACCTACTAGCTAATACGCCGCATCCTCATCTC CTAGCGGTCCAAACGGACCTTTTAACACCTTCTGATGCCATAATGGTGTCGTATGCGGTATTAGCAGTC GTTTCCAACTGTTATCCCCCACTAAGAGGTAGATTAGATACGTGTTACTCACCCGTTCGCCACTGGGTG CAAGCACCCCGTTCGACTTGCATGTATTAGGCATGCCGCCAGCG
ITS region	TTTCTATGGAGTAATACATAGATATATATAAGCTGAGTGAATATTAATAGACAGCTTTTGGCTTGTATC TTGGTTCTATCTAGTTTTCAGAGATTGTTTTAGTTTGTTTACAAACTTGACAATTTCTGAAAAGAAATC GTTTTTGAAAACTGAATATTAGATGAATTAGACAATTTTATTTTTCTTACGTTTCAATATAAACGTAAT AGTAACTTAAAAAGTTTTACTAAGAAAACTAAAATTTCAATTTTATATAGATTATCTATATTTGCAAT
23S rRNA	TATGATTAATCTCATAACATCTCTAACTGAATACATAGGTTATTGAAGGGAACCTAGGGAACTGAAAC ATCTTAGTACCTAGAGGAAAAGAAAACGAATGTGATTCTGTCAGTAGCGGCGAGCGAAAGCGGAACA GGCCAAACCGGTCTACGGGCCGGGGTTGTAGGACTCTTGTTAGAGTTACAAAATTAGTGTATAGCAGA AACTGTTGGGAAACAGTGGCATAGAGTGTGATACCCACGTATGCGAAATGCACTAATCTCGAGAGAG TATCCTGAGTACGGCGGGGCACGTGAAACCCTGTCGGAATCTACCCAGACCACTGGGTAAGCCTAAAT ACTACCATTTTACCGATAGTGAACCAGTACCGTGAGGGAAAGGTGAAAAGTACCCCGTGAGGGGAGT GAAATAGTACCTGAAACCATATGCTTACAAGAAGTCAGAGCCCGTTAATGGGTGATGGCGTGCTTTTT GTAGAAAGAGCCGGCGAGTTACGATATCATGCAAGGTTAAGTGGAATACACGGAGCCGCAGTGAAAG CGAGCCTTAATAGGGCGTTCAGTATGATGTCGTAGACACGAAACCAGGTGATCTAGCCATGAGCAGGT TGAAGTTAGGGTAAAACCTAATGGAGGACCGAACCGACGTTCGTTGAAATGACCGCGGATGACTTGT GGCTAGGGGTGAAATTCCAATCGAACCTGGAGATAGCTAGTTCTCCCCGATATAGCTTTAAGGCTAGC GTCGAGGTTAGGATTATGGAGGTAGAGCTCTGAATGTATGATGGCCCCACCTAGGGGTACTGATTACA ATTAAACTACGAATGCCATAATTTCATACTCGGCAGTCAGAACATGGGTGATAAGGTCCATGCTCGTG AGGGAAACAGCCCAGATCAACAACTAAGGTCCCTAAATTTATGCTAAGTGTGTAAGGATGTGGGAGT GCATAGACAGCTAGGATGTTGGCTCAGAAGCAGCCATCATTTAAAGAGTGCGTAACAGCTCACTAGTC GAGTGCTCCTGCGCCGAAAATGTACCGGGGCTTAAGCATAGTACCGAAGTTTTGGATTTACAGCAATG TAAGTGGTAGGGGAGCGTTCTAACTGTAATGAAGTTAAACCGTGAGGATTAGTGGAACGGTTAGAAG TGATTATGCCGGCATGAGTAACGTTTGGGAGTGAGAATCTCCCATGCCGTTTGACCAAGGTTTCCTGG GCAAGGTTCGTCCACCCAGGGTTAGTCAGGACCTAAGGCGAGGCCGAAAGGCGTAGTCGATGGACAA CAGGTTGATATTCCTGTACCACC
rpoB	CCATTTGGTACTCTTAAAGAGTTATCTTTAACATTTCTTGATTTTTCACCAAAGATTGCATGTAATAGTT TATCTTCAGGAGAAAGTTGTGTTTGAGATTTTGGTGTTACTTTACCAACTAAGATATCTCCAACTTTAA CTTCTGAACCAATTGCTACAATACCGTCTTCATCAAGATTTTTCTTAATAGCCTCAGAAACATTTGGAA TATCTCTTGTTATTTCTTCTTGACCTTGTTTTGTTTGTCTTCTTTCGATAGTATATTCATCAATATGAATT GAAGTGAATCTGTCATCGATTACAACTCTTTCTGAAACAATAACTGCATCTTCATAGTTATATCCATTT CAAGTTGTAAAAGCAACAACTACATTTTGACCAAGAGCAAGTTCACCTTTTTCCATTGAAGGACCATC GGCTAATATATCTCTTTTTTTAACTTTATCTCCAACTTTAACAATTGGGATGTGAGTAATTGCAGTTCCA TTATTTGATCTATTAAAATCATTTAAATCATAATTTTTAATACCATTTTTTTGTTCTACAACAATTCTTTT AGAATCAACATATTTTACAACTCCACCTTCAGTTGCAACAATTGCATCTCCTGAGTCTCTAGCAGCTTC AAATTCAACACCAGTTCCAACAACTGGAGATTCTGGATCAATTAAAGGTACGGCTTGACGTTGCATAT TGGCACCCATAAGCGCACGGTTGGCATCATCATTTTCCAAGAATGGAATACATGATGTAGCAATTGAA ACAATTTGTTTAGGTGAAAGTCCACATAATCAACATCTGATGAGTTAACCATAATGTCATCACCTCTGT ATCTTGCGATAACTTGGTCATCAATAATTGTTCCATCTTCTGCAATTTTAATATTTGCTTGTGCTACAAC ATATTCTTTTTCTTTATCAGCTGTTAAATATTCGTATTGATCTAAAATTACTTTTTTATTTTTAACTTTTC TATATGGTGTTTCAATAAATCCATATTCATTTATTTTTGCATA

Appendix B: *Spiroplasma* isolates/species with identical DNA sequences that were collapsed for phylogenetic analyses.

16S rRNA	ITS	23S rRNA	*rpoB*	16S & 23S rRNA	16S & 23S rRNA; ITS & *rpoB*	16 rRNA, ITS & *rpoB*
S. lineolae BARC 4900 BARC 4903 GSU5366 GSU5363	***S. lineolae*** BARC 4900 BARC 4903 GSU5363 GSU5366	**BARC 4900** GSU5366 GSU4886 BARC 4903		**BARC 4900** GSU5366 BARC 4903		
S. tabanidicola GSU4886						
GSU5400 GSU5360	***S. gladitoris*** BARC 4689 GSU4980 GSU5400 GSU5360	**BARC 4689** GSU4980 **GSU5400** BARC 5360	***S. gladitoris*** BARC 4689			
S. litorale GSU5382 GSU5405 GSU5443 GSU5853 GSU5865	**GSU5353** GSU5382 GSU5443 GSU5853 GSU5861 GSU5865	**GSU5353** GSU5443 GSU5853 GSU5865 GSU5861		**GSU5853** GSU5443 GSU5865		
S. helicoides GSU5420 GSU5441 GSU5446H GSU5867	***S. helicoides*** GSU5373 GSU5420 GSU5441 GSU5867	**GSU5450** GSU5446h GSU5420				
GSU5858 GSU5529 GSU5603 GSU5508	**GSU5858** GSU5529 GSU5603 GSU5508	**GSU5858** GSU5603 GSU5529	**GSU5858** GSU5529 GSU5603 GSU5508	**GSU5858** GSU5603 GSU5529	**GSU5858** GSU5529 GSU5603 GSU5508	**GSU5858** GSU5529 GSU5603
	BARC 4908 BARC 4806	**BARC 4908** BARC 4906	**BARC 4908** BARC 4906			
GSU5478 GSU5510	**GSU5478** GSU5510	**GSU5478** GSU5510	**GSU 5478** GSU5510	**GSU5478** GSU5510	**GSU5478** GSU5510	**GSU5478** GSU5510