Molecular phylogeny and a taxonomic proposal for the genus *Streptococcus*

F. Póntigo[1], M. Moraga[1,2] and S.V. Flores[1]

[1]Laboratorio de Antropología, Departamento de Antropología, Facultad de Ciencias Sociales, Universidad de Chile, Santiago, Chile
[2]Programa de Genética Humana, Instituto de Ciencias Biomédicas, Universidad de Chile, Santiago, Chile

Corresponding author: S.V. Flores
E-mail: sfloresc@uchile.cl

Genet. Mol. Res. 14 (3): 10905-10918 (2015)
Received February 4, 2015
Accepted June 17, 2015
Published September 21, 2015
DOI http://dx.doi.org/10.4238/2015.September.21.1

ABSTRACT. Alternative phylogenies for the genus *Streptococcus* have been proposed due to uncertainty about the among-species group relationships. Here, we performed a phylogenetic analysis of the genus *Streptococcus*, considering all the species groups and also the genomic data accumulated by other studies. Seventy-five species were subjected to a Bayesian phylogenetic analysis using sequences from eight genes (16S rRNA, *rpoB*, *sodA*, *tuf*, *rnpB*, *gyrB*, *dnaJ*, and *recN*). On the basis of our results, we propose a new Phylogeny for the genus, with special emphasis on the inter-species group level. This new phylogeny differs from those suggested previously. From topological and evolutionary distance criteria, we propose that gordonii, pluranimalium, and sobrinus should be considered as new species groups, in addition to the currently recognized groups of mutans, bovis, pyogenic, suis, mitis, and salivarius.

Key words: Bayesian analysis; Phylogeny; *Streptococcus*; Species group

INTRODUCTION

The genus *Streptococcus* is a diverse lineage belonging to the lactic acid group of bacteria. Current taxonomy places this genus, as well as the genus *Lactococcus*, within the family *Streptococcaceae* in the order Lactobaillales (Facklam, 2002). These organisms are Gram-positive, spherical, and catalase-negative, and many are facultative anaerobes (Montes and García-Arenzana, 2007; Lal et al., 2011).

Recent molecular systematic studies based on 16S rRNA sequences have provided evidence to subdivide *Streptococcus* into six "species groups": pyogenes, anginosus, mitis, salivarius, bovis, and mutans (Kawamura et al., 1995; Facklam, 2002; Rodicio and Mendoza, 2004; Montes and García-Arenzana, 2007; Lal et al., 2011).

In addition to the striking phylogenetic diversification, the ecological features of the members of the *Streptococcus* genus encompass a myriad of environments and life styles. While some species are well known for their clinical importance as pathogens, others have been characterized as members of a range of normal microbiomes in different anatomical structures in animals and humans. This distribution pattern explains their regular occurrence in contaminated biological samples (Montes and García-Arenzana, 2007; Glazunova et al., 2010; Boggs et al., 2012).

The features of pathogenicity associated with the genus *Streptococcus* are also diverse: meningitis, pneumonia, endocarditis, fasciitis, and dental caries are among the better known conditions (Glazunova et al., 2010). Normal human reservoirs of *Streptococcus* include different compartments of the oral cavity and skin, and the respiratory, digestive, gastric, and urinary tracts (Hardie and Whiley, 1997). However, the complete home range for most species of the genus is largely uncertain since this knowledge depends on sampling strategies that are not normally focused on revealing species habitats.

Systematic studies have been performed in the genus. Kawamura et al. (1995) carried out a pioneer study that proposed the current general phylogenetic organization for the genus *Streptococcus*. In that study, 16S rRNA gene sequences from 28 distinct species were analyzed leading to a hypothesis of six species groups that is still generally accepted.

Simmon et al. (2008) examined the viridans group using 16S rRNA, *tuf*, and *rpoB* gene sequences from 22 species and subspecies to evaluate phylogenetic relationships within the genus in a study of samples from patients affected by endocarditis. They demonstrated horizontal gene transfer and different rates of molecular substitutions among DNA fragments in the genus, making phylogenetic inferences a challenging subject for this lineage.

Furthermore, other studies have evaluated the phylogenetic certainty of different gene fragments. For example, Täpp et al. (2003) analyzed the gene *rnpB* in 50 species, in order to obtain phylogenetic information from a source other than 16S rRNA. Similarly, Itoh et al. (2006) proposed the *dnaJ* and *gyrB* genes as good phylogenetic markers, and analyzed 45 species and subspecies. Additionally, the *recN* gene has been shown to render statistically confident clades in an extensive analysis that included 60 species (Glazunova et al., 2010).

To date, the phylogenetic relationships among the known species groups have not been resolved. For example, Itoh et al. (2006) concluded that disagreements between the trees obtained from three different gene fragments could be explained either by horizontal gene transfer or different mutation rates. This also can be seen in the phylogenetic results of Simmon et al. (2008) who described phylogenetic discordances that could be explained

by horizontal gene transfer. In both studies, the analyses were performed using phylogenetic distances; the effects of different phylogenetic methods were not evaluated. Glazunova et al. (2006) and Täpp et al. (2003) used diverse genes and different phylogenetic approaches (distances and Bayesian analysis, respectively) and produced two alternative phylogenetic hypotheses (Figure 1).

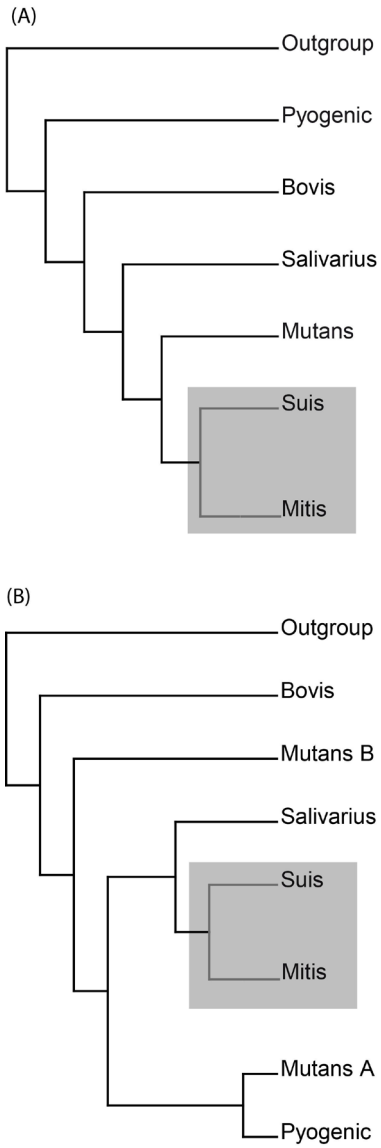

Figure 1. Phylogenetic hypothesis by (**A**) Glazunova et al. (2006, Figure 1), based on fragment sequences of *recN* gene; and (**B**) Tapp et al. (2003; Figure 3), using sequences of 16S r RNA and *rnpB*.

As shown in Figure 1, the species groups suis and mitis are in sister clades in both topologies. This is the only convergence between these phylogenetic hypotheses, as other relationships remained unclear.

Using the available genomic databases it is possible to incorporate all the gene fragment data that has been reported into a single phylogenetic analysis. This approach will enable the development of a new systematic work on this genus, covering all the known species groups. In this article, we attempt to create an extensive phylogeny for the genus *Streptococcus*, and contrast our proposed phylogeny with previously published hypotheses.

MATERIAL AND METHODS

Samples

DNA sequences from the genes *sodA*, *tuf*, *rpoB*, 16s rRNA, *rnpB*, *dnaJ*, *gyrB*, and *recN* were retrieved from GenBank (www.ncbi.nlm.nih.gov/genbank; Table 1). Seventy-five species and subspecies of *Streptococcus* were analyzed. *Lactococcus lactis* subsp *lactis* and *Lactococcus lactis* subsp *cremoris* were selected as outgroups, since this genus is one of the four members of the family Streptococcaceae. Although a primary analysis showed that *Pilibacter* is the sister clade of *Streptococcus*, the lack of data for this genus did not allow robust analyses to be performed (Figure 2).

Molecular phylogenetic analyses

Nucleotide sequences were aligned using the software BioEdit V. 7.0.9.0 (Hall, 1999). The alignments were checked and re-edited by eye. In order to avoid the effect of saturation of the phylogenetic signal, especially due to synonymous substitutions, amino acid sequences were also used for the six protein coding genes (*sodA*, *tuf*, *rpoB*, *dnaJ*, *gyrB*, and *recN*). The non-coding genes 16s rRNA and *rnpB* were included in both analyses of nucleotide sequences and amino acid derived sequences. Both nucleotide and amino acid derived sequences were concatenated in the same super-matrix. Absent data were coded as missing data and incorporated in subsequent analyses. This "super-matrix" approach has been shown to perform better than eliminating data, or using non-concatenated matrices, especially when Bayesian methods are used for phylogenetic inference (Wiens and Moen, 2008).

Models of molecular evolution for nucleotide sequences were selected using the Akaike information criterion test (AIC), as implemented in the software Modeltest (Posada and Crandall, 1998) and MrModeltest (Nylander, 2004). Models for amino acid sequences were selected using the PROTTEST v. 3.0 software (Abascal et al., 2005).

Bayesian inference on the complete super matrix was performed using Mr. Bayes v. 3.0b4, specifying four million generations and four independent chains. The resulting tree identified seven subclades that were used as the basis for subsequent analyses (Figure 3 and Table 2).

In order to solve phylogenetic relationships among species groups, a new matrix was built sampling two species belonging to each of the groups. This matrix, called the "mini-matrix", was used for Bayesian inference using the same conditions and model substitutions as described above for the complete super-matrix. Phylogenetic concordance among the different DNA sequences was assessed by the "homogeneity test" (Farris et al., 1994, 1995) as implemented in Paup, in order to evaluate the phylogenetic compatibility among gene fragments inside the super-matrix.

Table 1. List of species, accession numbers and size (bp) of each gene fragment.

Species	*sodA*		*tuf*		*rpoB*		*recN*	
	bp	Accession No.	bp	Accession No.	bp	Accession No.	bp	Accession No.
S. mutans	612	AE014133.2	1197	AE014133.2	3105	AP010655.1	1249	EU917289.1
S. agalactiae	609	AE009948.1	1197	AL766847.1	3105	AE009948.1	1249	EU917242.1
S. acidominimus	435	Z95892.1	761	AY266992.1	691	AF535181.1	1249	EU917241.1
S. anginosus	453	FJ712177.1	826	AF276257.1	3105	AF535183.1	1249	EU917248.1
S. alactolyticus	435	AJ297185.1	-	-	680	DQ232445.1	1249	EU917226.1
S. australis	435	GU907529.1	761	EU156907.1	680	DQ132983.1	-	-
S. caballi	399	EF364100.1	-	-	-	-	-	-
S. canis	435	Z99175.1	761	EU156908.1	680	DQ232488.1	-	-
S. castoreus	435	EU003820.1	-	-	680	EU003817.1	1249	EU917312.1
S. criceti	435	Z95898.1	821	AF276260.1	680	DQ232485.1	1249	EU917257.1
S. cristatus	435	GU907530.1	761	AF276261.2	2061	AF194513.1, AF194514.1, AB199920.1	1249	EU917258.1
S. devriesei	435	DQ232544.1	-	-	680	DQ232446.1	1249	EU917306.1
S. didelphis	435	DQ232545.1	-	-	680	DQ232447.1	1249	EU917259.1
S. downei	435	Z95899.1	792	AF276262.1	680	DQ132984.1	-	-
S. dysgalactiae subsp *equisimilis*	486	AP010935.1	1215	AP010935.1	3105	AP010935.1	1249	EU917227.1
S. constellatus	453	FJ712176.1	821	AF276259.1	680	FJ712146.1	1093	FJ712113.1
S. intermedius	453	FJ712187.1	799	AF276267.1	-	-	1249	EU917282.1
S. entericus	435	DQ232547.1	-	-	680	DQ232448.1	1249	EU917228.1
S. equi subsp. equi	486	FM204883.1	1100	FM204883.1	3105	FM204883.1	1249	EU917229.1
S. equinus	435	AJ297213.1	792	AF276258.1	3105	AF535187.1	1249	EU917252.1
S. ferus	435	DQ132986.1	792	AF276265.1	686	AY770000.1	1249	EU917265.1
S. gallinaceus	435	DQ232548.1	-	-	680	DQ232503.1	1246	EU917266.1
S. gallolyticus	486	FN597254.1	1100	FN597254.1	3105	FN597254.1	-	-
S. gordonii str. Challis substr. CH1	486	CP000725.1	1100	CP000725.1	3105	CP000725.1	-	-
S. halichoeri	435	DQ232571.1	-	-	680	DQ232471.1	1249	EU917277.1
S. henryi	399	EF364099.1	-	-	-	-	-	-
S. hyointestinalis	435	DQ232550.1	-	-	680	DQ232449.1	1249	EU917231.1
S. hyovaginalis	435	DQ232551.1	-	-	680	DQ232450.1	1249	EU917232.1
S. infantarius subsp *infantarius*	435	AJ306980.1	761	EU156915.1	672	EU420169.1	-	-
S. infantis	432	GU907532.1	761	EU156917.1	680	DQ232482.1	1249	EU917280.1
S. iniae	486	EU661272.1	761	EU156918.1	680	DQ232493.1	1249	EU917281.1
S. luteciae	435	AJ297212.1	-	-	-	-		
S. lutetiensis	426	GU991740.1	761	EU156916.1	680	DQ232480.1	1249	EU917279.1
S. macacae	435	DQ232553.1	792	AF276268.1	680	DQ232452.1	1249	EU917285.1
S. macedonicus	435	AJ297187.1|	761	EU156913.1	693	AY315156.1	1249	HE613569.1
S. marimammalium	435	EU003821.1	-	-	659	EU003818.1	1249	EU917313.1
S. massiliensis	451	AY769999.1	761	EU156919.1	691	AY769998.1	1249	EU917311.1
S. minor	435	DQ232572.1	-	-	680	DQ232472.1	1249	EU917286.1
S. mitis	486	FN568063.1	1100	FN568063.1	3105	FN568063.1	1249	EU917233.1
S. oligofermentans	435	DQ232554.1	761	EU156926.1	680	DQ232508.1	1249	EU917290.1
S. oralis	435	DQ232576.1	764	AF276270.1	2658	AY695496.1, AB199956.1	1249	EU917234.1
S. orisratti	435	DQ232555.1	-	-	680	DQ232453.1	1249	EU917235.1
S. ovis	435	DQ232556.1	-	-	680	DQ232454.1	1249	EU917236.1
S. parasanguinis	435	GU907536.1	789	AF276271.1	-	-	1249	CP002843.1
S. parauberis	435	AJ544723.1	761	EU156937.1	680	DQ232455.1	1212	EU917293.1
S. pasteurianus	435	DQ232583.1	761	EU156914.1	680	DQ232505.1	1249	EU917273.1
S. peroris	435	DQ232541.1	761	EU156938.1	680	DQ232483.1	1249	EU917294.1
S. phocae	435	AJ547799.1	-	-	680	DQ232456.1	1249	EU917237.1
S. pluranimalium	435	DQ232557.1	761	EU156939.1	680	DQ232457.1	1249	EU917295.1
S. plurextorum	356	AM774231.1	-	-	687	AM774232.1	-	-
S. pneumoniae	486	FM211187.1	1100	FM211187.1	3105	FM211187.1	1249	EU917238.1
S. porcinus	435	Z99177.1	-	-	680	DQ232486.1	1249	EU917296.1
S. pseudopneumoniae	390	AB200048.1	761	EU156954.1	680	EU003819.1	1249	EU917314.1
S. pyogenes	486	CP000262.1	1100	AE009949.1	3105	CP000261.1	1249	EU917297.1
S. ratti	435	DQ232559.1	792	AF276272.1	680	DQ232484.1	1249	EU917298.1
S. salivarius	432	GU907538.1	792	AF276273.1	691	AF535169.1	1249	EU917239.1
S. sanguinis	486	CP000387.1	1100	CP000387.1	2922	CP000387.1	1249	EU917299.1

Continued on next page

Table 1. Continued.

Species	*sodA*		*tuf*		*rpoB*		*recN*	
	bp	Accession No.	bp	Accession No.	bp	Accession No.	bp	Accession No.
S. sinensis	435	DQ232560.1	761	EU156962.1	680	DQ232458.1	1249	EU917305.1
S. sobrinus	435	Z95919.1	780	AF276275.1	680	DQ232489.1	-	-
S. suis	486	CP000408.1	1100	CP000408.1	3102	AM946016.1	1249	CP002570.1
S. thermophilus	486	CP000024.1	1100	CP000024.1	3105	CP000024.1	1249	EU917300.1
S. thoraltensis	435	DQ232562.1	-	-	680	DQ232461.1	1249	EU917303.1
S. uberis	486	AM946015.1	1100	AM946015.1	3105	AM946015.1	1249	EU917309.1
S. urinalis	435	DQ232561.1	761	EU156963.1	680	DQ232460.1	1249	EU917240.1
S. vestibularis	436	AY687381.1	781	AF276277.1	686	AY687377.1	1249	EU917304.1
L. lactis subsp *lactis*	481	CP001834.1	1084	CP001834.1	3105	CP001834.1	-	-
L. lactis subsp *cremoris*	467	CP000425.1	1084	CP000425.1	3105	CP000425.1	-	-
Species	*16SrRNA*		*rnpB*		*dnaJ*		*gyrB*	
	bp	Accession No.	bp	Accession No.	bp	Accession No.	bp	Accession No.
S. mutans	1542	AB294730.1	380	AJ511678.1	1134	AE014133.2	1953	AP010655.1
S. agalactiae	1537	AF459432.1	374	AJ511673.1	1140	AL766843.1	1953	AE009948.1
S. acidominimus	1334	X58301.1	373	AJ511681.1	967	AB239170.1	900	AB238609.1
S. anginosus	1559	NR_041722.1	370	AJ511731.1	971	AB238697.1	900	AB236189.1
S. alactolyticus	1501	EU728776.1	376	AJ511706.1	-	-	-	-
S. australis	1471	NR_036936.1	-	-	648	GU907541.1	458	EU003771.1
S. caballi	1430	NR_044190.1	-	-	-	-	-	-
S. canis	1482	EU075058.1	374	AJ511684.1	969	AB238698.1	900	AB236191.1
S. castoreus	1508	NR_042215.1	-	-	-	-	458	EU003814.1
S. criceti	1530	EU483241.1	-	-	964	AB238699.1	900	AB238610.1
S. cristatus	1533	NR_042771.1	373	AJ511700.1	896	AB238700.1	900	AB238611.1
S. dentapri	1550	AB469560.1	-	-	-	-	-	-
S. dentirousetti	1543	NR_041460.1	-	-	-	-	-	-
S. devriesei	1471	EU483245.1	-	-	-	-	457	EU003742.1
S. didelphis	1506	AF176107.1	-	-	914	AB238810.1	457	EU003790.1
S. downei	1542	NR_042774.1	376	AJ511699.1	-	-	900	AB238618.1
S. dysgalactiae subsp *equisimilis*	1492	AP010935.1	374	CP002215.1	1137	AP010935.1	1953	AP010935.1
S. constellatus	1508	AY277942.1	370	AJ511742.1	964	AB238808.1	509	AB441109.1
S. intermedius	1531	GU470908.1	374	AJ511743.1	547	AB441138.1	1947	AB562520.1
S. entericus	1335	NR_025500.1	-	-	-	-	-	-
S. equi subsp. equi	1495	FM204883.1	374	FM204883.1	1137	FM204883.1	1953	FM204883.1
S. equinus	1539	AB362710.1	373	AJ511745.1	916	AB238704.1	-	-
S. ferus	1540	AB259060.1	366	AJ511705.1	902	AB238705.1	903	AB238616.1
S. gallinaceus	1502	NR_025453.1	-	-	-	-	458	EU003785.1
S. gallolyticus	1535	AF323911.1	375	AJ511683.1	1143	AP012053.1	1953	AP012053.1
S. gordonii str. Challis substr. CH1	1499	CP000725.1	-	-	1134	CP000725.1	1950	CP000725.1
S. halichoeri	1506	NR_029025.1	-	-	-	-	458	EU003791.1
S. henryi	1430	NR_044189.1	-	-	-	-	-	-
S. hyointestinalis	1504	EU728763.1	372	AJ511696.1	1014	AB238709.1	900	AB238621.1
S. hyovaginalis	928	EF151158.1/ DQ118670.1	376	AJ512493.1	974	AB238710.1	900	AB238622.1
S. ictaluri	1471	DQ462421.1	-	-	-	-	-	-
S. infantarius subsp *infantarius*	1494	EU163504.1	371	AJ511688.1	1008	AB238812.1	458	EU003810.1
S. infantis	1468	NR_042928.1	387	AJ511687.1	835	AB238711.1	900	AB238623.1
S. iniae	1536	NR_025148.1	378	AJ511708.1	905	AB238712.1	900	GU324259.1
S. luteciae	1461	NR_042051.1	-	-	-	-	-	-
S. lutetiensis	1501	JN713319.1	371	AJ511709.1	-	-	458	EU003729.1
S. macacae	1542	NR_042775.1	373	AJ511702.1	1051	AB238713.1	900	AB238625.1
S. macedonicus	1538	AF459431.1	374	AJ511677.1	-	-	458	HE613569.1
S. marimammalium	1500	NR_025630.1	-	-	-	-	458	EU003815.1
S. massiliensis	1470	NR_043173.1	-	-	-	-	458	EU003813.1
S. merionis	1354	NR_042553.1	-	-	-	-	-	-
S. minor	1497	AY232833.1	-	-	-	-	458	EU003795.1
S. mitis	1540	AY518677.1	387	AJ511694.1	1140	FN568063.1	1947	FN568063.1
S. oligofermentans	1510	NR_029052.1	-	-	-	-	458	EU003768.1
S. oralis	1470	DQ232535.1	387	AJ511698.1	1137	FR720602.1	1947	FR720602.1
S. orisratti	1474	EU075064.1	374	AJ511692.1	1043	AB238813.1	458	EU003798.1

Continued on next page

Table 1. Continued.

Species	*16SrRNA*		*rnpB*		*dnaJ*		*gyrB*	
	bp	Accession No.	bp	Accession No.	bp	Accession No.	bp	Accession No.
S. orisuis	1553	NR_041055.1	-	-	-	-	-	-
S. ovis	1377	NR_026471.1	-	-	-	-	458	EU003799.1
S. parasanguinis	1513	NR_024842.1	373	AJ511704.1	974	AB238798.1	900	CP002843.1
S. parauberis	1541	AF284579.2	374	CP002471.1	890	AB238799.1	900	AB238630.1
S. pasteurianus	1511	EU163495.1	-	-	-	-	458	AP012054.1
S. peroris	1329	EU156772.1	387	AJ511690.1	890	AB238800.1	900	AB238631.1
S. phocae	1510	HM032023.1	371	AJ511670.1	935	AB238801.1	900	AB238632.1
S. pluranimalium	1552	Y18026.1	375	AJ511697.1	-	-	458	EU003801.1
S. plurextorum	1451	NR_042649.1	-	-	-	-	-	-
S. pneumoniae	1511	AM157442.1	386	CP002121.1	1137	AE005672.3	1947	CP000918.1
S. porcinus	1496	NR_024634.1	369	AJ511675.1	-	-	900	AB175052.1
S. pseudopneumoniae	1468	CP002925.1	-	-	-	-	458	EU003816.1
S. pseudoporcinus	1491	DQ303207.1	-	-	-	-	-	-
S. pyogenes	1518	FJ662846.1	374	CP003121.1	1137	AE009949.1	1953	CP000262.1
S. ratti	1444	NR_025516.1	373	AJ511671.1	896	AB238804.1	900	AB238635.1
S. salivarius	1543	AF459433.1	373	FR873482.1	621	GU907550.1	900	AB238640.1
S. sanguinis	1510	AY691542.1	373	AJ511682.1	1134	CP000387.1	1950	CP000387.1
S. seminale	1275	AB370977.1	-	-	-	-	-	-
S. sinensis	1512	AF432855.1	-	-	-	-	458	EU003809.1
S. sobrinus	1551	AB294731.1	373	AJ511707.1	-	-	900	AB238641.1
S. suis	1536	AF009497.1	370	AJ511674.1	1137	CP002633.1	1953	CP002641.1
S. thermophilus	1544	EF990662.1	373	AJ511712.1	1134	FR875178.1	1953	CP002340.1
S. thoraltensis	1519	NR_026368.1	-	-	885	AB238806.1	900	AB238637.1
S. uberis	1501	NR_040820.1	372	AJ511693.1	1134	AM946015.1	1953	AM946015.1
S. urinalis	1476	DQ303194.1	373	AJ511680.1	-	-	458	EU003763.1
S. ursoris	1549	AB501126.1	-	-	-	-	-	-
S. vestibularis	1553	NR_042777.1	373	AJ511724.1	-	-	900	AB238643.1
L. lactis subsp *Lactis*	1539	CP001834.1	-	-	1140	CP001834.1	1956	CP001834.1
L. lactis subsp *cremoris*	1539	CP000425.1	-	-	1088	CP000425.1	1956	CP000425.1
Lactovum miscens	1456	NR_042140						
Pilibacter termitis	1432	NR_042949						

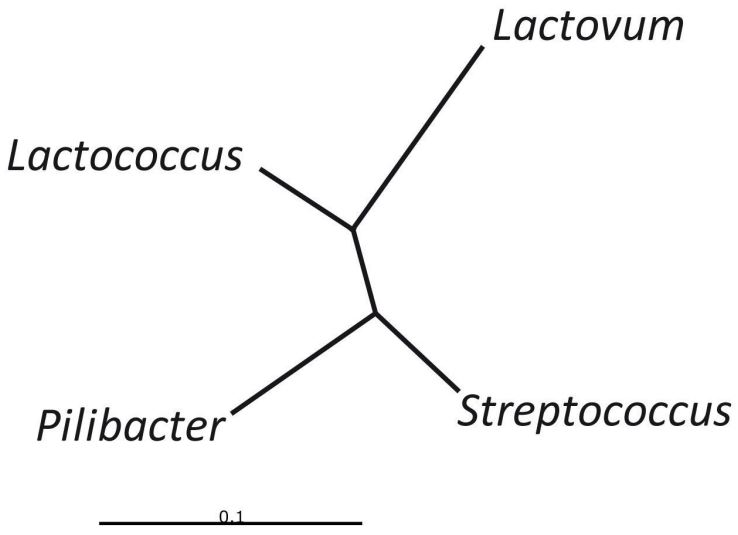

Figure 2. Unrooted tree obtained from a Bayesian analysis of representative species of the family Streptococcaceae, inferred from their 16sR RNA sequences (see details in Table 1).

RESULTS AND DISCUSSION

Seven major clades were detected in our analyses; these corresponded to the conventional species groups, except for the group mitis that was split into two non-monophyletic clades (Figure 3). In general, phylogenies inferred either from nucleotide or amino acid derived sequences did not differ either on major-clade relationships or statistical confidence (i.e., posterior probabilities).

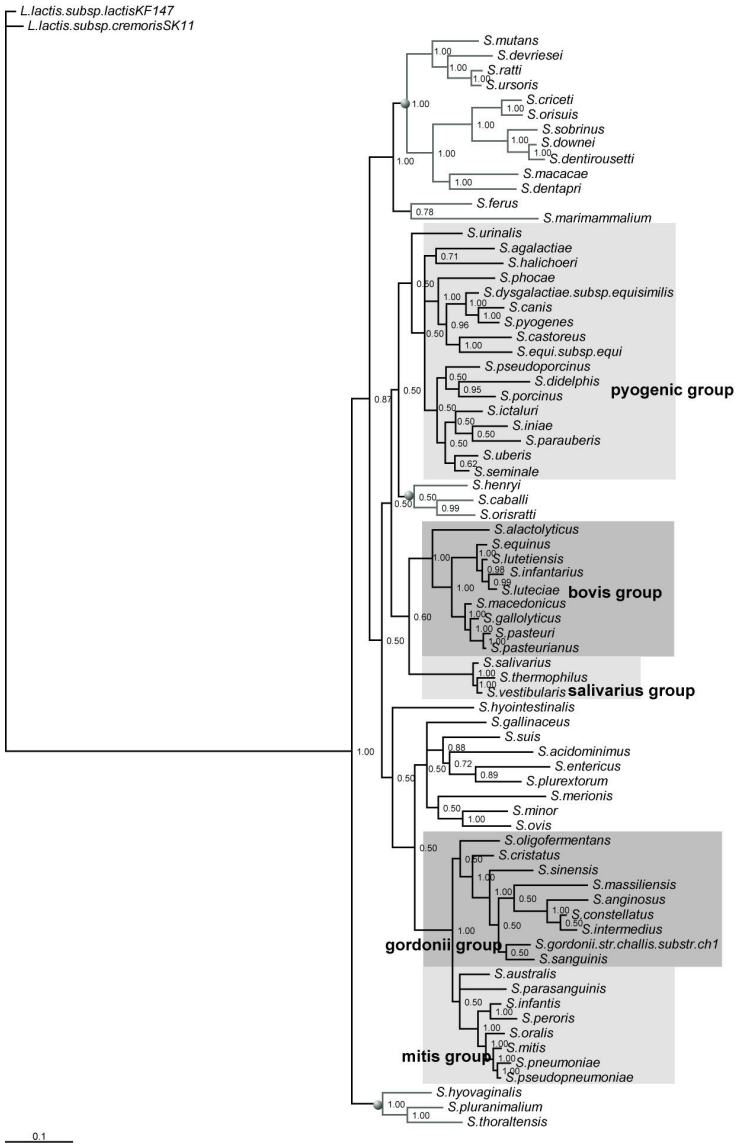

Figure 3. Majority consensus tree obtained by a Bayesian analysis of proteic (sodA, tuf, rpoB, dnaJ, gyrB) and nucleotidic (16sr RNA, rnpB) of the entire matrix.

Table 2. Species of *Streptococcus*, and outgroup species, analyzed in this study.

Mutans group	*S. ratti*, *S. ursoris*, *S. devriesei*, *S. mutans*, *S. macacae*, *S. ferus*, *S. dentapri*, *S. downei*, *S. dentirousetti*, *S. sobrinus*, *S. dentirousetti*, *S. sobrinus*, *S. criceti*, *S. orisuis*, *S. merionis*, *S. caballi*, *S. henryi*, *S. orisratti*, *S. pluranimalium*, *S. thoraltensis*, *S. hyovaginalis*
Pyogenic group	*S. dysgalactiae*, *S. pyogenes*, *S. canis*, *S. castoreus*, *S. equi* subsp *equi*, *S. halichoeri*, *S. phocae*, *S. porcinus*, *S. pseudoporcinus*, *S. didelphis*, *S. uberis*, *S. seminale*, *S. iniae*, *S. ictaluri*, *S. urinalis*, *S. parauberis*, *S. marimammalium*, *S. agalactiae*
Bovis group	*S. equinus*, *S. lutetiensis*, *S. luteciae*, *S. infantarius*, *S. gallolyticus*, *S. pasteurianus*, *S. macedonicus*, *S. alactolyticus*
Suis group	*S. entericus*, *S. plurextorum*, *S. suis*, *S. acidominimus*, *S. minor*, *S. ovis*, *S. gallinaceus*
Mitis group A	*S. constellatus*, *S. intermedius*, *S. anginosus*, *S. massiliensis*, *S. cristatus*, *S. sinensis*, *S. gordonii*, *S. sanguinis*
Mitis group B	*S. pneumoniae*, *S. pseudopneumoniae*, *S. mitis*, *S. oligofermentans*, *S. infantis*, *S. peroris*, *S. oralis*, *S. australis*, *S. parasanguinis*
Salivarius group	*S. vestibularis*, *S. salivarius*, *S. thermophilus*
Outgroup	*Lactococcus lactis*, *L. cremoris*

With regard to the phylogenetic relationships within species groups, it is notable that the mitis group (Kawamura et al., 1995) remained monophyletic but, in agreement with the results of Boggs et al. (2012) (Figure 3), was clearly subdivided into two clades. The suis group was basal to the sister clades.

Here we propose a taxonomic revision, since our results indicate that divergence between the two major clades subdividing the group mitis is in the range of divergence for other species groups (within-mean distances for the "gordonii" and "mitis" clades were 0.039 and 0.044, respectively; between-mean distance was 0.073). Therefore, a new species group was proposed and named gordonii; the group included the species *S. constellatus*, *S. intermedius*, *S. anginosus* [referred to as species group anginosus by Kawamura et al. (1995)], plus *S. massiliensis*, *S. cristatus*, *S. sinensis*, *S. gordonii*, and *S. sanguinis*. Additionally, the group mitis was retained, and included the species *S. australis*, *S. parasanguinis*, *S. oralis*, *S. peroris*, *S. infantis*, *S. oligofermentans*, *S. mitis*, *S. pseudopneumoniae*, and *S. pneumoniae.*

We suggest that gordonii is the basal group of the genus, followed by mitis, suis and finally pyogenic (Figure 3). The most recent divergence was the separation of the sister species groups salivarius and bovis. All the species groups, except mutans, resulted in monophyletic groups (Figure 3). *S. ferus* was basal to the entire genus, and not inside the mutans group as previously reported. Similarly, *S. entericus* and *S. plurextorum* (suis group), *S. marimammalium* and *S. agalactiae* (pyogenic group) were not placed in their previously described groups. The remaining species were clustered inside their formerly ascribed groups.

The diverse analyses performed here clearly show that the majority of the species groups had a stable topology. By changing the parameters of the gene sequences, i.e., using amino acid derived or nucleotide sequences, or by constraining clades, most of the species groups did not change their species composition. In the major part of the non-constrained analysis, the mutans species group resulted in a polyphyletic taxon, compressing at least three different lineages (Figure 3).

However, a homogeneity analysis revealed incompatibilities among gene sequences within the mutans species group; these incompatibilities might be due to horizontal gene transfer (Table 3) (Abascal et al., 2005; Itoh et al., 2006; Simmon et al., 2008; Boggs et al., 2012). More critical incompatibilities, detected by pairwise analysis, allowed us to identify large genetic discordances between different loci. We took these results into account in subsequent analyses, avoiding incompatibilities and generating different strategies in order to infer phylogenetic relationships within each species group and among groups.

Table 3. Homogeneity test results. P values for all the pairwise combinations of gene and protein fragments are shown for each species group and for the "mini matrix" set of species.

	Mutans group	Pyogenic group	Bovis group	Suis group	Mitis group A	Mitis group B	Mini matrix
All	0.01	0.01	0.27	0.88	0.01	0.01	0.01
sodA x tuf	0.26	0.12	0.56	0.99	0.14	0.01	0.04
sodA x rpoB	0.80	0.42	0.94	0.96	0.71	0.01	0.23
sodA x 16sr	0.02	0.01	0.30	0.53	0.17	0.01	0.03
sodA x rnpB	0.01	0.58	0.05	0.97	0.42	0.01	0.16
sodA x dnaJ	0.01	0.61	0.93	0.99	0.68	0.39	0.01
sodA x gyrB	0.27	0.04	0.88	0.38	0.62	0.03	0.93
sodA x recN	0.01	0.43	1	0.06	0.31	0.01	0.31
tuf x rpoB	0.41	0.37	1	1	0.51	0.01	0.02
tuf x 16sr	0.28	0.45	0.38	0.98	0.07	0.02	0.04
tuf x rnpB	0.01	0.99	0.56	1	0.72	0.01	0.01
tuf x dnaJ	0.01	0.83	1	1	0.77	0.50	0.01
tuf x gyrB	0.87	0.40	0.98	1	0.03	0.05	0.50
tuf x recN	0.01	0.83	0.89	0.97	0.26	0.01	0.23
rpoB x 16sr	0.41	0.08	0.74	0.66	0.51	0.07	0.01
rpoB x rnpB	0.17	0.91	1	1	0.35	0.31	0.01
rpoB x dnaJ	0.03	0.78	1	1	0.21	0.28	0.03
rpoB x gyrB	0.99	0.60	1	0.72	0.49	0.65	0.47
rpoB x recN	0.22	0.90	0.06	0.12	0.19	0.79	0.07
16sr x rnpB	0.05	0.16	0.49	0.98	0.01	0.05	0.09
16sr x dnaJ	0.10	0.29	0.92	1	0.69	0.14	0.01
16sr x gyrB	0.04	0.01	0.76	0.85	0.54	0.24	0.59
16sr x recN	0.01	0.01	0.88	0.41	0.68	0.01	0.33
rnpB x dnaJ	0.29	1	1	1	0.01	0.94	0.01
rnpB x gyrB	0.02	0.42	0.87	1	0.35	0.39	0.07
rnpB x recN	0.01	0.85	0.03	0.95	0.06	1	0.01
dnaJ x gyrB	0.01	0.67	1	1	0.09	0.31	0.02
dnaJ x recN	0.01	0.80	1	0.99	0.01	0.56	0.01
gyrB x recN	0.01	0.77	0.98	1	0.25	0.59	0.87

For example, with regard to the mutans group, only *sodA*, *tuf*, *rpoB*, and *gyrB* could be jointly analyzed, because only these genes shared high homogeneity scores (Table 3). Further analysis of this clade considered the entire matrix with the concatenated sequences of *soda*, *rpoB*, and *gyrB*. These gene sequences were used in a new sampling strategy, which included all the species from the mutans group, plus two species representing each of the other species groups.

Next, we focused on the unsolved part of the tree: the mutans group and its relationship with salivarius, bovis, and pyogenic groups, using the same gene sequences and parameters as described above. This analysis reinforced the conclusion that the mutans group was not a monophyletic taxon. Rather, it appeared to be constituted by at least three lineages, namely mutans, sobrinus, and pluranimalium. This Bayesian tree resolved most of the relationships between these species (Figure 4).

Subsequently, we focused on unresolved parts of the phylogeny, namely, the relationship of *S. macacae* to the mutans and pyogenic groups, and the relationship between the pluranimalium and sobrinus groups. Homogeneity tests allowed us to select pairs of gene sequences to analyze for this part of the phylogeny (Table 4).

Bayesian analyses of *tuf-recN rpoB-recN* gene sequences placed *S. macacae* inside the mutans group, with *S. ferus*, *S. mutans*, *S. devriesei*, and *S. ratti* (Figure 5a). Bayesian trees of *sodA-rpoB*, 16S rRNA-*rnpB*, and *sodA-recN* gene sequences identified these two groups as sister clades (Figure 5b). Nonetheless, the phylogenetic relationships of *S. caballi* and *S. orisratti* remained unsolved since insufficient nucleotide data were present in the databases. Moreover, there was a lack of phylogenetic signal due to the low rate of nucleotide substitution (i.e., 16S rRNA) and phylogenetic incompatibility between the gene sequences (i.e. between 16S rRNA and *sodA*).

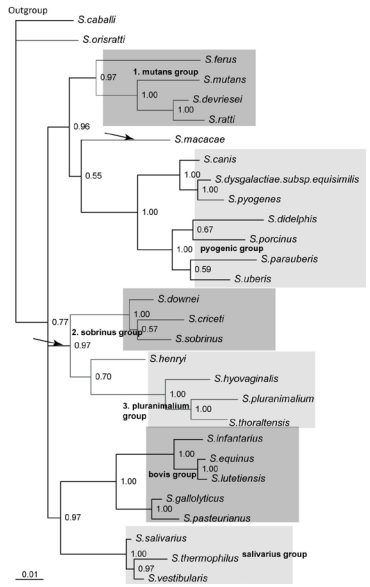

Figure 4. Majority consensus tree obtained from a Bayesian analysis of concatenated sequences of a subtree of *Streptococcus* using 3 proteic fragments (sodA, rpoB, gyrB). Arrows indicate unsolved points of the topology. Species groups suis, gordonii and mitis are excluded from this subtree. The models of protein evolution were WAG (sodA) and Rtrev (rpoB, recN).

Table 4. Homogeneity test results. P values for all the pairwise combinations of gene and protein fragments are shown for each one of the two pairs of minigroups derived from the mutans group.

	Clade A	Clade B
sodA x tuf	0.56	0.14
sodA x rpoB	0.98	0.87
sodA x 16s	0.72	0.01
sodA x rnpB	0.94	0.01
sodA x dnaJ	0.97	0.13
sodA x gyrB	0.54	0.31
sodA x recN	0.98	0.56
tuf x rpoB	0.92	0.68
tuf x 16s	0.75	0.63
tuf x rnpB	1	0.14
tuf x dnaJ	1	0.55
tuf x gyrB	0.90	0.33
tuf x recN	1	0.95
rpoB x 16s	0.95	0.02
rpoB x rnpB	1	0.04
rpoB x dnaJ	1	0.57
rpoB x gyrB	1	0.56
rpoB x recN	0.85	0.94
16s x rnpB	0.98	0.08
16s x dnaJ	0.96	0.06
16s x gyrB	0.54	0.01
16s x recN	0.92	0.01
rnpB x dnaJ	1	0.08
rnpB x gyrB	0.87	0.01
rnpB x recN	1	0.57
dnaJ x gyrB	1	0.06
dnaJ x recN	1	0.78
gyrB x recN	0.94	0.79

Clade A = *Streptococcus pluranimalium*, *S. thoraltensis*, *S. hyovaginalis*, *S. henryi*, and *S. criceti*, *S. sobrinus*, *S. downei*. Clade B = *S. devriesei*, *S. ratti*, *S. mutans*, *S. ferus*, *S. macacae*, *S. canis*, *S. pyogenes*, *S. dysgalactiae*, *S. didelphis*, *S. porcinus*, and *S. uberis*.

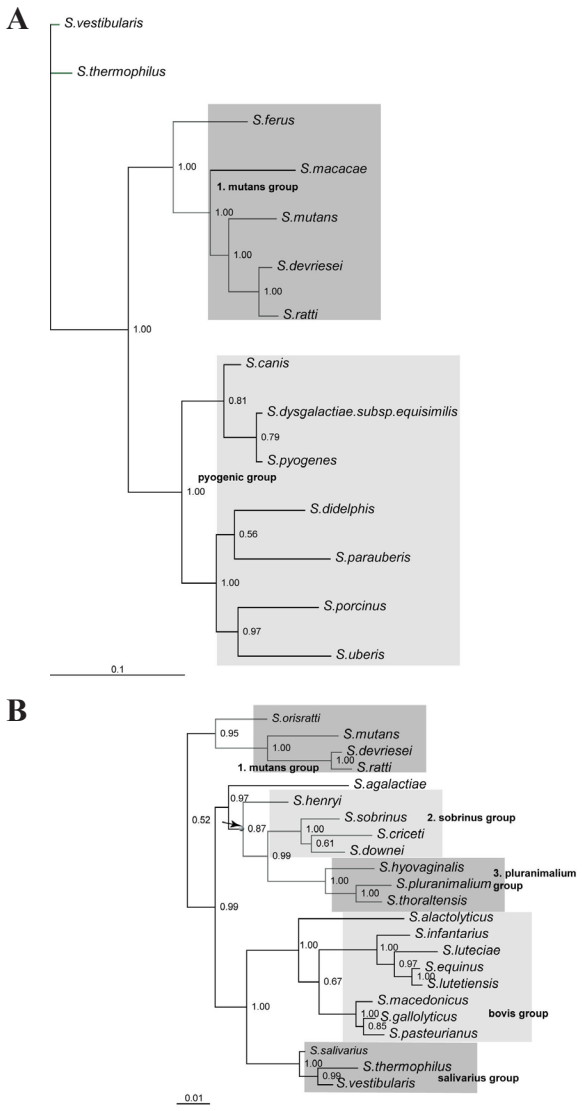

Figure 5. Majority consensus tree obtained of a Bayesian analysis of concatenated sequences. Those matrices were built using (**A**) 2 aminoacidic fragments (rpoB, recN), in order to solve the phylogenetic position of *Streptococcus macacae*; (**B**) 2 aminoacidic fragments (sodA, recN) to infer the relationship between sobrinus and pluranimalium groups. The models of protein evolution were WAG (sodA), Rtrev (rpoB), and Jones (recN).

We performed a phylogenetic analysis of a mini-matrix comprising two species from each of the nine groups (mutans, sobrinus, pluranimalium, salivarius, bovis, pyogenic, gordonii, mitis, and suis; Figure 6). The results showed that, in contrast to previous studies, the suis, gordonii, and mitis groups were basal clades. In contrast, sobrinus and pluranimalium were sister clades and closely related to the salivarius group. This phylogeny is shown in Figure 7, and represents our hypothesis on the phylogenetic relationships among the species groups of *Streptococcus*.

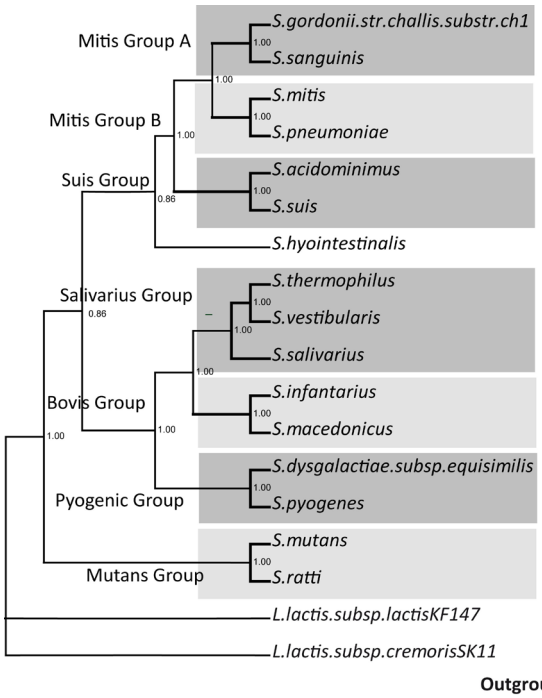

Figure 6. Majority consensus tree obtained by a Bayesian analysis of proteic (sodA, rpoB, and gyrB) matrix of the summarized tree containing two members of each group, plus the outgroup.

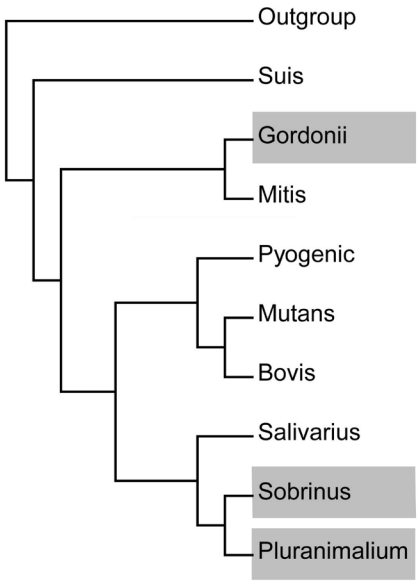

Figure 7. Hypothesis suggested by this study, summarizing the results obtained from the complete super-matrix and the mini-matrix analyses.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by grants FONDECYT #1110461, FONDECYT #1140544, VID-Enlace #70219 and Programa de Investigación Asociativa Conicyt, Anillo #ACT-096.

REFERENCES

Abascal F, Zardoya R and Posada D (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-2105.

Boggs JM, South AH and Hughes AL (2012). Phylogenetic analysis supports horizontal gene transfer of L-amino acid oxidase gene in *Streptococcus oligofermentans*. *Infect. Genet. Evol.* 12: 1005-1009.

Facklam R (2002). What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin. Microbiol. Rev.* 15: 613-630.

Farris JS, Källersjö M, Kluge AG and Bult C (1994). Testing significance of incongruence. *Cladistics* 10: 315-319.

Farris JS, Källersjö M, Kluge AG and Bult C (1995). Constructing a significance test for incongruence. *Syst. Biol.* 44: 570-572.

Glazunova OO, Raoult D and Roux V (2006). *Streptococcus massiliensis* sp. nov., isolated from a patient blood culture. *Int. J. Syst. Evol. Microbiol.* 56: 1127-1131.

Glazunova OO, Raoult D and Roux V (2010). Partial recN gene sequencing: a new tool for identification and phylogeny within the genus *Streptococcus*. *Int. J. Syst. Evol. Microbiol.* 60: 2140-2148.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser*. 41: 95-98.

Hardie JM and Whiley RA (1997). Classification and overview of the genera *Streptococcus* and *Enterococcus*. *Soc. Appl. Bacteriol. Symp. Ser.* 83: 1S-11S.

Itoh Y, Kawamura Y, Kasai H, Shah MM, et al. (2006). dnaJ and gyrB gene sequence relationship among species and strains of genus *Streptococcus*. *Syst. Appl. Microbiol.* 29: 368-374.

Kawamura Y, Hou XG, Sultana F, Miura H, et al. (1995). Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int. J. Syst. Bacteriol.* 45: 406-408.

Lal D, Verma M and Lal R (2011). Exploring internal features of 16S rRNA gene for identification of clinically relevant species of the genus *Streptococcus*. *Ann. Clin. Microbiol. Antimicrob.* 10: 28.

Montes M and García-Arenzana JM (2007). Género *Streptococcus*: una revisión práctica para el laboratorio de microbiología. *Enferm. Infecc. Microbiol. Clin.* 25: 14-20.

Nylander JAA (2004). MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre. Uppsala University.

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

Rodicio M and Mendoza MC (2004). Identificación bacteriana mediante secuenciación del ARNr 16S: fundamento, metodología y aplicaciones en microbiología clínica. *Enferm. Infecc. Microbiol. Clin.* 22: 238-245.

Simmon KE, Hall L, Woods CW, Marco F, et al. (2008). Phylogenetic analysis of viridans group streptococci causing endocarditis. *J. Clin. Microbiol.* 46: 3087-3090.

Täpp J, Thollesson M and Herrmann B (2003). Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, rnpB. *Int. J. Syst. Evol. Microbiol.* 53: 1861-1871.

Wiens JJ and Moen DS (2008). Missing data and the accuracy of Bayesian phylogenetics. *J. Syst. Evol.* 46: 307-314.