Genetics and Molecular Research

Streptococcus massiliensis in the human mouth: a phylogenetic approach for the inference of bacterial habitats

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

[1]Departamento de Antropología, FACSO, Universidad de Chile, Santiago, Chile
[2]Programa de Genética Humana, ICBM, Universidad de Chile, Santiago, Chile

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

Genet. Mol. Res. 14 (4): 19184-19190 (2015)
Received August 15, 2015
Accepted October 26, 2015
Published December 29, 2015
DOI http://dx.doi.org/10.4238/2015.December.29.28

ABSTRACT. *Streptococcus* is a diverse bacterial lineage. Species of this genus occupy a myriad of environments inside humans and other animals. Despite the elucidation of several of these habitats, many remain to be identified. Here, we explore a methodological approach to reveal unknown bacterial environments. Specifically, we inferred the phylogeny of the Mitis group by analyzing the sequences of eight genes. In addition, information regarding habitat use of species belonging to this group was obtained from the scientific literature. The oral cavity emerged as a potential, previously unknown, environment of *Streptococcus massiliensis*. This phylogeny-based prediction was confirmed by species-specific polymerase chain reaction (PCR) amplification. We propose employing a similar approach, i.e., use of bibliographic data and molecular phylogenetics as predictive methods, and species-specific PCR as confirmation, in order to reveal other unknown habitats in further bacterial taxa.

Key words: *Streptococcus massiliensis*; Microbiome; Phylogeny; Mitis group, RecN

INTRODUCTION

The genus *Streptococcus* is a diverse lineage consisting of gram-positive, spherical, catalase-negative organisms, including many facultative anaerobes (Montes and García-Arenzana, 2007). This group inhabits a broad range of environments, not only as pathogens, but also as part of the normal microbiome of humans and other animals.

Based on 16S rRNA gene phylogenies, this genus has been separated into six species groups (Kawamura et al., 1995). Among them, the Mitis group comprises diverse species of significant clinical importance, including *Streptococcus pneumoniae* and *S. sanguinis*, the pathogens responsible for pneumonia and endocarditis, respectively. Current opinion suggests that this group contains at least 17 species (Kawamura et al., 1995). Recently, *S. massiliensis*, a new species discovered in a sample of human blood, was added (Glazunova et al., 2006).

In the current study, we analyzed the Mitis group phylogeny and, in combination with published data of known sample origins, used this information to infer unknown habitats. Finally, we tested for the presence of *S. massiliensis* in human mouths, since our preliminary phylogenetic analysis suggested the oral cavity as a previously unidentified habitat for this species.

MATERIAL AND METHODS

Phylogenetic analyses

Sequences of the following genes were retrieved from GenBank and used to infer the Mitis group phylogeny: superoxide dismutase (*sodA;* Poyart et al., 2002; Whatmore and Whiley, 2002), elongation factor Tu, RNA polymerase subunit β (*tuf* and *rpoB*, respectively; Simmon et al., 2008), 16S rRNA (Kawamura et al., 1995), RNase P subunit β (*rnpB*; Täpp et al., 2003), heat shock protein 40, DNA gyrase subunit B (*dnaJ* and *gyrB*, respectively; Itoh et al., 2006), and recombination protein N (*recN*; Glazunova et al., 2010; Table 1). The 17 species composing the Mitis group (Kawamura et al., 1995, 1999; Glazunova et al., 2006; Boggs et al., 2012) were included in our analyses. Their close phylogenetic relationship was corroborated in a new analysis using concatenated fragments of *recN* and *gyrB* genes (data not shown). Two outgroups belonging to the genus *Lactococcus* were used, namely *L. lactis* subsp *lactis* and *L. lactis* subsp *cremoris*.

Sequences were aligned using BioEdit version 7.0.9.0 (Hall, 1999) and alignments were checked and re-edited by eye. Nucleotide sequences were concatenated into a single supermatrix. Missing data were coded as such and incorporated into subsequent analyses. This “supermatrix” strategy has been shown to perform better than strategies involving the elimination of data or the use of non-concatenated matrices, particularly when Bayesian methods are used for phylogenetic inference (Wiens and Moen, 2008).

Molecular evolution models for nucleotide sequences were selected using the Akaike information criterion test, as implemented in the programs Modeltest (Posada and Crandall, 1998) and MrModeltest (Nylander, 2004). Bayesian inference was performed on the supermatrix using MrBayes version 3.0b4, specifying 4,000,000 generations and four independent chains.

Subsequently, known habitats for each of the 17 species were identified by searching the literature (Table 2) and the information gathered was mapped onto the phylogenetic tree.

Table 1. List of *Streptococcus* and *Lactococcus* sequences retrieved from GenBank used in the phylogenetic analysis.

Species	*sodA*		*tuf*		*rpoB*		*recN*	
	bp	Accession No.	bp	Accession No.	bp	Accession No.	bp	Accession No.
S. anginosus	453	FJ712177.1	826	AF276257.1	3105	AF535183.1	1249	EU917248.1
S. cristatus	435	GU907530.1	761	AF276261.2	2061	AF194513.1 AF194514.1 AB199920.1	1249	EU917258.1
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. gordonii	486	CP000725.1	1100	CP000725.1	3105	CP000725.1		
S. massiliensis	451	AY769999.1	761	EU156919.1	691	AY769998.1	1249	EU917311.1
S. sanguinis	486	CP000387.1	1100	CP000387.1	2922	CP000387.1	1249	EU917299.1
S. sinensis	435	DQ232560.1	761	EU156962.1	680	DQ232458.1	1249	EU917305.1
S. australis	435	GU907529.1	761	EU156907.1	680	DQ132983.1	-	-
S. infantis	432	GU907532.1	761	EU156917.1	680	DQ232482.1	1249	EU917280.1
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. parasanguinis	435	GU907536.1	789	AF276271.1	-	-	1249	CP002843.1
S. peroris	435	DQ232541.1	761	EU156938.1	680	DQ232483.1	1249	EU917294.1
S. pneumoniae	486	FM211187.1	1100	FM211187.1	3105	FM211187.1	1249	EU917238.1
S. pseudopneumoniae	390	AB200048.1	761	EU156954.1	680	EU003819.1	1249	EU917314.1
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	-	-
	16S rRNA			rnpB		dnaJ		gyrB
S. anginosus	1559	NR_041722.1	370	AJ511731.1	971	AB238697.1	900	AB236189.1
S. australis	1471	NR_036936.1	-	-	648	GU907541.1	458	EU003771.1
S. cristatus	1533	NR_042771.1	373	AJ511700.1	896	AB238700.1	900	AB238611.1
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. gordonii	1499	CP000725.1	-	-	1134	CP000725.1	1950	CP000725.1
S. infantis	1468	NR_042928.1	387	AJ511687.1	835	AB238711.1	900	AB238623.1
S. massiliensis	1470	NR_043173.1	-	-	-	-	458	EU003813.1
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. parasanguinis	1513	NR_024842.1	373	AJ511704.1	974	AB238798.1	900	CP002843.1
S. peroris	1329	EU156772.1	387	AJ511690.1	890	AB238800.1	900	AB238631.1
S. pneumoniae	1511	AM157442.1	386	CP002121.1	1137	AE005672.3	1947	CP000918.1
S. pseudopneumoniae	1468	CP002925.1	-	-	-	-	458	EU003816.1
S. sanguinis	1510	AY691542.1	373	AJ511682.1	1134	CP000387.1	1950	CP000387.1
S. sinensis	1512	AF432855.1	-	-	-	-	458	EU003809.1
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

Samples

Saliva was collected from 29 volunteers. All participants signed an informed consent form, which prevents use of the data collected for purposes other than this research, and protects their confidentiality. This document was reviewed and approved by the Ethics Committee for Research in Social Science and Humanities (CEDEA) of the Faculty of Philosophy at the University of Chile. Two or three DNA extractions were carried out for each sample of saliva (Quinque et al., 2006).

Polymerase chain reaction (PCR)

Primers used to amplify a fragment of the *recN* gene were designed using Primer3 Input (Untergrasser et al., 2012) and a recent sequence for *S. massiliensis* (GenBank accession No.

EU917311.1). This gene fragment was selected based on the high variability that it demonstrated in other species of the Mitis group in a preliminary analysis (Table 3).

Table 2. Habitats occupied by the *Streptococcus* and *Lactococcus* species included in this study.

Species	Habitat
S. anginosus	Oral cavity, gastrointestinal and urogenital tracts
S. australis	Oral cavity
S. cristatus	Oral cavity
S. constellatus	Oropharynx, respiratory tract
S. intermedius	Oral cavity, respiratory tract
S. gordonii	Oral cavity
S. infantis	Oral cavity, upper respiratory tract (humans)
S. massiliensis	Blood sample
S. mitis	Oral cavity
S. oligofermentans	Oral cavity
S. oralis	Oral cavity
S. parasanguinis	Oral cavity (humans)
S. peroris	Oral cavity (humans)
S. pneumoniae	Nasopharynx, oral cavity (humans), nasopharynx (pigs, cats, horses, dogs)
S. pseudopneumoniae	Phlegm (humans)
S. sanguinis	Oral cavity
S. sinensis	Oral cavity
L. lactis subsp *lactis*	Dairy products
L. lactis subsp *cremoris*	Dairy products

Table 3. Proportion of variable sites in different gene fragments from species belonging to the *Mitis* species group.

Gene fragment	Variability	Proportion of variable sites
rpoB	223/680	0.3
sodA	220/486	0.45
recN	669/1258	0.53

Primer sequences and their respective melting temperatures were as follows: MassiF, 5'-CGA CTT GGA CGC GAT TAG CG-3', 56.27°C; and MassiR, 5'-TAC CTG CTC ATT GCC CTC GCG GTT-3', 62.53°C. The amplified sequence fragment comprised nucleotide positions 634 to 1111 of the *recN* gene.

In order to determinate the optimal annealing temperature for amplification, a thermal gradient PCR was performed, ranging from 55° to 65°C in 12 intervals over 35 cycles, using two replicates per sample and a negative control for each temperature.

Seven successfully amplified products were submitted for direct sequencing to Macrogen Inc. (Seoul, Korea). BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches were performed in order to confirm the species-specific amplification expected from the primer design.

RESULTS

Phylogenetic analysis

Bayesian analysis resulted in a tree with high posterior probabilities for the majority of its internal nodes (Figure 1). Two well-defined subclades, namely the Mitis and Gordonii groups, are shown. Mapping of habitats onto the phylogeny revealed that the majority of the species under investigation inhabit the oral cavity. The species of particular interest in this study, *S. massiliensis*, was found to be basal to the subclade [(*S. constellatus*, *S. intermedius*), *S. anginosus*].

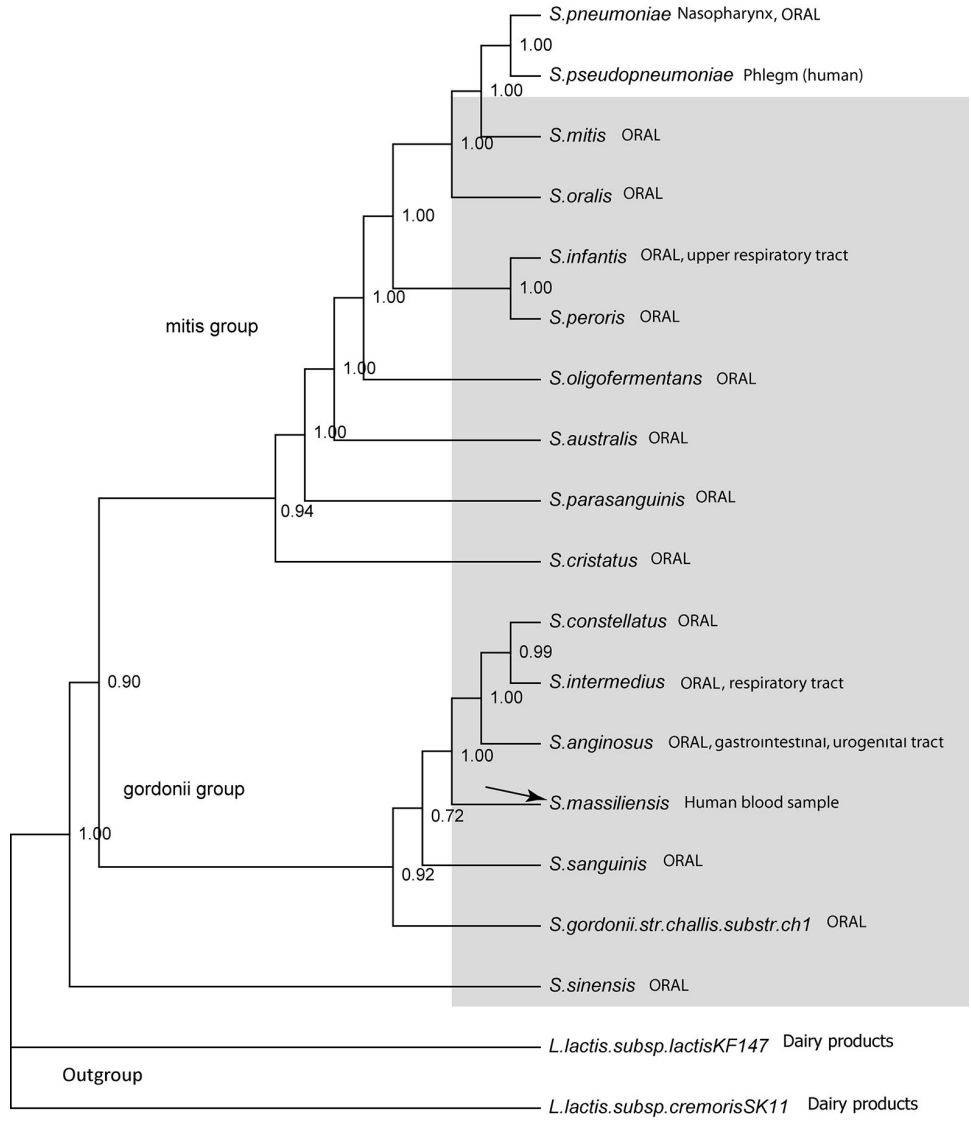

Figure 1. Majority-rule consensus tree obtained from a Bayesian analysis utilizing concatenated sequences from fragments of eight genes (16S rRNA, *rnpB*, *rpoB*, *dnaJ*, *gyrB*, *sodA*, *tuf*, and *recN*). Seventeen *Streptococcus* species were included, along with two *Lactococcus* species as outgroups. In addition, the habitat occupied by each species was mapped onto the tree according to information obtained from the literature. The general time-reversible + gamma + invariant sites nucleotide substitution model was used.

PCR

PCR was carried out on several samples using various numbers of cycles, as shown in Table 4. Optimal results were achieved with 40 cycles and an annealing temperature of 64.7°C.

Table 4. Results of polymerase chain reaction amplifications using different numbers of cycles.

	35 cycles	40 cycles	45 cycles	50 cycles
Samples with successfully amplified products	4	11	4	4
Samples with non-specific products	0	1	1	8
Samples with non-amplified product	25	13	14	13

DNA sequencing and analysis

As shown in Figure 2, no differences were detected between the seven sequences obtained and the reference sequence (GenBank accession No. EU917311.1), indicating that the amplicons derived from *S. massiliensis*.

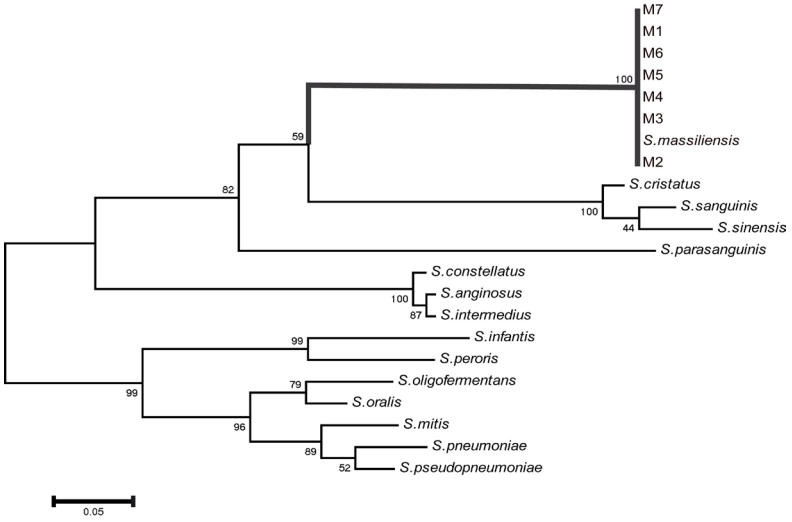

Figure 2. Electrophoresis of *Streptococcus massiliensis* polymerase chain reaction products (35 cycles, 75 V, 75 min). Lanes and their respective samples are labelled.

DISCUSSION

In the present study, we demonstrated that *S. massiliensis*, a *Streptococcus* species recently discovered in human blood, is a regular inhabitant of the human oral cavity, as predicted from habitat-use data and a phylogeny of the Mitis species group. Similar approaches could be employed to reveal unknown habitats occupied by other lineages of bacteria and *Streptococcus* species.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the following grants: FONDECYT (#1110461), FONDECYT (#1140544); and Programa de Investigación Asociativa Conicyt, Anillo (#ACT-096).

REFERENCES

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.

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. Nucleic Acids Symposium Series, 41: 95-98.

Handley P, Coykendall A, Beighton D, Hardie JM, et al. (1991). *Streptococcus crista* sp. nov., a Viridans *Streptococcus* with tufted fibrils, isolated from the human oral cavity and throat. *Int. J. Syst. Bacteriol*. 41: 543-547.

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.

Kawamura Y, Whiley RA, Shu SE, Ezaki T, et al. (1999). Genetic approaches to the identification of the mitis group within the genus *Streptococcus. Microbiology* 145: 2605-2613.

Montes M and García-Arenzana JM (2007). Género Streptococcus: una revisión práctica para el laboratorio de microbiología. *Enferm. Infecc. Microbiol. Clin.* 25 (Suppl 3): 14-20.

Nylander JAA (2004). 'MrModeltest v2', Computer Program, Evolutionary Biology Centre, Uppsala University.

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

Poyart C, Quesne G and Trieu-Cuot P (2002). Taxonomic dissection of the *Streptococcus bovis* group by analysis of manganese-dependent superoxide dismutase gene (*sodA*) sequences: reclassification of '*Streptococcus infantarius* subsp. *coli'* as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype 11.2 as *Streptococcus pasteurianus* sp. nov. *Int. J. Syst. Evol. Microbiol*. 52: 1247-1255.

Quinque D, Kittler R, Kayser M, Stoneking M, et al. (2006). Evaluation of saliva as a source of human DNA for population and association studies. *Anal. Biochem*. 353: 272-277.

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.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, et al. (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* 40: e115.

Whatmore AM and Whiley RA (2002). Re-evaluation of the taxonomic position of *Streptococcus ferus*. *Int. J. Syst. Evol. Microbiol.* 52: 1783-1787.

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