Infection, Genetics and Evolution 12 (2012) 1005–1009

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Phylogenetic analysis supports horizontal gene transfer of L-amino acid oxidase gene in *Streptococcus oligofermentans*

Joseph M. Boggs, April H. South, Austin L. Hughes *

Department of Biological Sciences, University of South Carolina, 715 Sumter St., Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Received 25 October 2011
Received in revised form 21 February 2012
Accepted 24 February 2012
Available online 3 March 2012

Keywords:
Horizontal gene transfer
L-Amino acid oxidase
Streptococcus oligofermentans

ABSTRACT

Phylogenetic analysis of 10 amino acid sequences from 19 *Streptococcus* species showed that *S. oligofermentans* clustered within the mitis group. However, the L-amino acid oxidase (LAAO) of *S. oligofermentans* showed a different clustering pattern from the other proteins analyzed implicating horizontal gene transfer (HGT) in the origin of the *S. oligofermentans LAAO* gene. LAAO of *S. oligofermentans* is known to confer ability to compete with other oral cavity bacteria, most notably *S. mutans*; therefore, the HGT event may have been important in extending the ecological niche occupied by this species, consistent with those of other studies suggesting that HGT can play a key role in enabling bacterial species to occupy new ecological niches.

1. Introduction

The genus *Streptococcus* contains several species of economic and medical importance, including species that are commensals or pathogens of man and domestic animals (Delorme et al., 2010; Glazunova et al., 2006; Matta et al., 2009). Among the most important impacts of *Streptococcus* on human health is their role in dental caries, which is one of the most prevalent bacterial diseases affecting human populations (Petersen, 2008; Petersen et al., 2005). The primary source of cariogenesis can be isolated to one microbial species, *S. mutans*, which was first described by Fitzgerald and Keyes (1960) as the etiological agent of cariogenesis. *S. mutans* is only one of many species among the oral flora which produce cariogenesis through acidogenesis. However, *S. mutans* displays the unique ability to bond to the tooth surface using a mechanism involving the protein coded by *spaP* (Ajdic et al., 2002). The ability to bond to an uncolonized tooth surface sets *S. mutans* apart as a gateway species. After *S. mutans* colonization more fastidious bacteria colonize through interspecies glucan interactions and further the demineralization of the tooth structure through acid production.

S. oligofermentans is a relatively novel species first described by Tong et al. in 2003. The species' importance revolves around its ability to inhibit *S. mutans* in the oral environment. Studies in the Chinese population have confirmed that *S. oligofermentans* in addition to being non-cariogenic, is also found on tooth surfaces free of *S. mutans*. *S. oligofermentans*' inhibitory properties towards *S. mutans* have been mapped to the *lox* gene, encoding L-amino acid oxidase gene (LAAO; Tong et al., 2007, 2008). A knockout of the *lox* gene in *S. oligofermentans* lost inhibitory properties towards *S. mutans* (Tong et al., 2008). This inhibitory capacity, along with the lowered tooth surface-binding ability of *S. oligofermentans* and lowered production of acids which cause hydroxyapatite demineralization in the teeth, suggest that *S. oligofermentans* may be useful in prophylactic measures against dental caries (Zhang et al., 2010).

The genus *Streptococcus* has been previously categorized based on phylogenies constructed using 16S rRNA into six groups: pyogenic, anginosus, mitis, salivarius, bovis, and mutans groups (Kawamura et al., 1995). Though widely used for bacterial phylogeny, 16s rRNA sequences are not always useful in the study of closely related species because of high sequence identity due to functional constraint. Thus, protein-coding genes have also been used, as illustrated by a phylogenetic analysis of the mitis group based on partial manganese-dependent superoxide dismutase gene (*sodA*) sequences (Kawamura et al., 1999).

In the present study, we reconstructed phylogenetic relationships of selected *Streptococcus* isolates based on the amino acid sequences encoded by *sodA* and nine other genes, including the gene encoding LAAO. Our goals were twofold: (1) to examine the evolutionary relationship between *S. oligofermentans* and the major previously identified groups of *Streptococcus*; and (2) to compare the phylogenies reconstructed using different genes in order to test for evidence of recombinational events in their evolutionary histories, as indicated by phylogenetic inconsistencies between genes.

2. Materials and methods

2.1. Amino acid sequences

S. oligofermentans amino acid sequences from the Genbank database were used in a BLAST homology search in order to

* Corresponding author. Tel.: +1 803 777 9186; fax: +1 803 777 4002.
E-mail address: austin@biol.sc.edu (A.L. Hughes).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2012.02.014

identify homologs in other *Streptococcus* species (Supplementary Table S1). In the analysis we used all amino acid sequences available in the database from the following taxa: *S. oligofermentans* and 18 other *Streptococcus* isolates (*S.* species 2 1 36FAA, *S. thermophilus*, *S. suis*, *S. sanguinis*, *S. salivarius*, *S. pyogenes*, *S. pneumoniae*, *S. parasanguinis*, *S. oralis*, *S. oral species*, *S. mutans*, *S. mitis*, *S.* M143, *S. gordonii*, *S. gallolyticus*, *S. dysgalactiae*, *S. bovis*, and *S. agalactiae*) and an outgroup within the family Streptococcaceae, *Lactococcus lactis*. Because multiple strains within species were available only in a few cases and because in *Streptococcus* within-species sequence differences are generally much lower than between-species differences (Pombert et al., 2009), for each protein we chose only a single sequence from each species, with preference for sequences from completely sequenced genomes.

The genes analyzed were the following: *PheS*, encoding phenylalanyl-tRNA synthase alpha subunit; *atpA* encoding ATP synthase F1 sector subunit alpha; *LAAO* encoding L-amino oxidase; *rpoB* encoding the DNA-directed RNA polymerase beta chain; *sodA* encoding superoxide dismutase; *groeEL* encoding the chaperonin groE; *gyrB* encoding DNA gyrase subunit B; recN encoding DNA repair protein recN; *atpD* encoding ATP synthase F1 sector subunit beta; and *EF-TU* encoding elongation factor TU (Supplementary Table S1). These genes are widely scattered throughout Streptococcus genomes. For example, in the sequenced genome of *S. gordonii* (NC_009785), the region including the 10 genes spans 1.2 megabases out of the 2.2 megabases in the genome.

2.2. Sequence alignment and phylogenetic reconstruction

Amino acid sequences were aligned using the CLUSTAL algorithm in MEGA 5 (Tamura et al., 2011). A Bayesian phylogenetic analysis of the concatenated alignment of all 10 amino acid sequences was conducted using the Mr. Bayes software (Huelsenbeck and Ronquist 2001) and the amino acid WAG + gamma model (Whelan and Goldman 2001), with four rate categories. The WAG + gamma model was chosen by the Mr. Bayes software as the most appropriate model for these data. Four chains were run for 1,700,000 generations with a sample taken every 100 generations producing a total of 34,002 trees separated equally into two files of which Bayesian posterior probabilities were inferred from 25,502 trees (12,751 per file). The average standard deviation of split frequencies was 0.005166.

We likewise conducted maximum likelihood (ML) phylogenetic analyses based on the WAG + gamma model with eight discrete rate categories, using the MEGA 5 software (Tamura et al. 2011). ML analyses were conducted for concatenated sequences; separately for each of the 10 proteins; and separately for the concatenated set, leaving out each protein in turn. Confidence of branching patterns within the ML phylogeny was assessed by bootstrapping (Felsenstein 1985); 1000 bootstrap samples were used. Most probable ancestral sequences were reconstructed by the ML method in MEGA 5.

2.3. Other statistical methods

We used the RDP3 program (Martin et al., 2010) to test for intragenic recombination. As a test for recombinational events involving entire genes, matrices of pairwise JTT + gamma distances amino acid distances were computed in MEGA 5, using shape parameters estimated by the ML method in MEGA 5 (Tamura et al., 2011). For each protein, the pairwise JTT + gamma distances were compared with that of the concatenated set of the nine remaining proteins. Plots of the pairwise amino acid distances for a given protein vs. those for the concatenated set of the nine other proteins were used to identify outliers possibly indicative of recombination.

A randomization test was used to test for significance of outliers. This randomization test was based on the regression of the distance in the protein of interest (Y) vs. that in the nine other proteins (X) and the computation of standardized residuals from that regression. Regression through the origin was used because it is expected that both X and Y would be zero immediately after divergence of two lineages. We created 1000 simulated samples by drawing randomly (with replacement) from the observed X and Y values. Each simulated sample contained as many bivariate data points (n) as the original sample. We computed the linear regression for each simulated sample and the absolute value of each standardized residual for each simulated sample.

To test whether a set of m standardized residuals from the original regression were significantly greater than expected, we compared the mean of the absolute value of those m standardized residuals with the mean of the absolute values of the remaining $n - m$ standardized residuals. We used as a test statistic the absolute difference (D) between the mean of the absolute values of the m standardized residuals in the set of interest and the mean of the absolute values of the remaining $n - m$ standardized residuals. The value of D in the actual data was compared with D values computed from groups of size m drawn at random from the simulated samples. For a two-tailed test, D in the actual data was considered significant at the α level if it was greater than the maximum D computed in $100(1 - \alpha)\%$ of the 1000 simulated samples.

3. Results

3.1. Phylogenetic analyses

The maximum likelihood phylogeny based on the concatenated set of all 10 amino acid sequences showed a cluster with 100% bootstrap support containing *S.* M143, *S. oralis*, *S. mitis, S. pneumoniae, S. oral species*, and *S. oligofermentans* (Fig. 1). With 77% bootstrap support, the latter group clustered with *S. sanguinis, S.* species 2 1 36FAA, and *S. gordonii* (Fig. 1). In addition, *S. parasanguinis* clustered with the above groups, and this 10-taxon cluster received 99% bootstrap support (Fig. 1). The latter cluster seemed to correspond to the mitis group as previously defined (Kawamura et al., 1995). The phylogeny constructed based on Bayesian methods

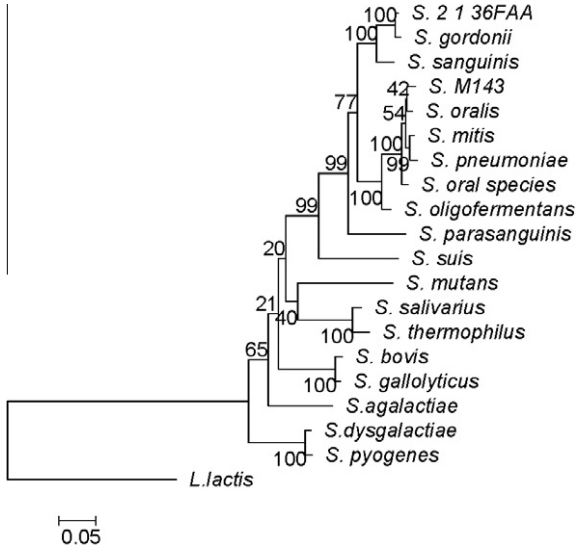

Fig. 1. ML tree based on 10 concatenated amino acid sequences (2561 aligned amino acid sites). Numbers on the branches represent percentage of bootstrap samples supporting the branch; only values ⩾50% are shown.