Abstract:

Streptococcus gallolyticus, formerly known as *S. bovis*, is part of the *Streptococcus bovis/Streptococcus equinus* complex (SBSEC) and is a normal inhabitant of the gastrointestinal tract in both humans and animals. This organism acts as both a commensal and an opportunistic pathogen across various host species. *S. gallolyticus* has attracted attention due to its association with infective endocarditis and its strong correlation with colorectal cancer in humans. In veterinary medicine, there have been spontaneous reports of systemic infections in animals, including swine, ruminants, and avian species. This study reviews the group's previous and current taxonomic classification and validates those approaches to phylogenetic analysis using Maximum likelihood and Bayesian analysis. This will address inconsistencies in (sub)species identification found in previous studies and offer an alternative approach. This new approach may offer a more accurate understanding of SBSEC evolution and help establish a standardized phylogenetic framework for the species.

Goals:

1. The phylogenetic classification of *Streptococcus gallolyticus* uses two well-known methods (Poyart et al., 2002 and Schlegel et al., 2003), applying 16S rRNA and *sodA* genes. The goal is to evaluate which method offers better phylogenetic resolution and classification
2. To validate the previous studies using Maximum Likelihood (ML) and Bayesian inference.

Hypotheses

1) H1: The sodA gene will yield higher phylogenetic resolution and stronger clade support than 16S rRNA for classifying S. gallolyticus strains across all tree-building methods.H3:
2) H2: The classification framework proposed by Schlegel et al. will result in phylogenetic trees that more accurately reflect known taxonomic relationships than the framework of Poyart et al., regardless of the gene or phylogenetic method used.
3) H3: Bayesian and maximum likelihood-based phylogenetic methods will produce more reliable and resolved phylogenetic trees than distance-based methods (Neighbor-Joining) for classifying S. gallolyticus strains.

Introduction

Method of classification from Poyart et al., 2002

1. The ***sodA*** sequences were analyzed with Perkin Elmer software (Sequence Analysis, Sequence Navigator, and AutoAssembler). Multiple alignments of the
2. The construction of the unrooted phylogenetic trees was performed with both the neighbor-joining with JC model
3. The reliability of the tree nodes was evaluated by calculating the percentage of 1000 bootstrap re-samplings that supported each topological element.

Method of classification from Schlegel et al., 2003

1. We determined the **16S rDNA** sequences of *S. infantarius* subsp *infantarius* HDP 90056T, S. *infantarius* subsp. *coli* HDP 90246T, *S. bovis* biotype II.2 HDP 90084, *S. equinus* HDP 89506T
2. We selected long (>1350 bp) and high-quality (<1% undetermined positions) sequences of type or reference strains. The length of the alignment was further limited to 1350 sites to reduce the weight of gaps and mismatches at the beginning or end of the sequences.
3. 16S rDNA data were performed with the megalign program from the DNAstar package. Sequences were aligned by using the clustal multiple-sequence method. A distance matrix was then computed using a Kimura model for nucleotide substitution.
4. Phylogenetic trees were generated from the distance matrices by using the neighbor-joining method.

Material and methods

#Sample: 20 strains of *S. gallolyticus*

#Prep for 16s

- Select long (>1350 bp) and high-quality (<1% undetermined positions) sequences and trim alignments to 1350 bp to reduce noise from sequence ends.
- Trimmed the long sequence from 350-400?

#Prep for soda

- Select the reference sodA and extract it

My methods:

Step 1: Apply Poyart et al.'s Method

- Use both 16S rRNA and sodA sequences from 40 S. gallolyticus strains.
- Perform multiple sequence alignment using MAFFT
- Build Neighbor-Joining (NJ) trees, unrooted using the phylip package (JC)
- Assess tree support with bootstrap analysis (1,000 replicates)

Step 2: Apply Schlegel et al.'s Method

- Perform alignment using mafft and generate a distance matrix using the Kimura model.
- Construct NJ trees for both 16S and sodA genes.
- Build Neighbor-Joining (NJ) trees, unrooted using the phylip package (Kimura)
- Assess tree support with bootstrap analysis (1,000 replicates)

Con: There might not been so much difference between the 2 models. We can compare which gene gets better support and whether the topology differs.

Step 3: initial comparison between the two

- Tree topology: Use Robinson-Foulds distance to measure differences.
- Bootstrap support: Compare average bootstrap percentages across methods.

- Clade consistency: Evaluate whether major clades appear consistently across trees.
- Rooting and resolution: Assess how well the tree separates lineages and whether it's well-resolved.
- Biological relevance: Compare results to known taxonomy or clinical patterns.
- "Face to face," aka "tanglegram," using Dendroscope if work.

Step 4: Add ML and Bayesian Phylogenies

- Maximum Likelihood (ML): RAxML for both genes using the GTR+G model (should try to do kimura to see the difference in the method of constructing tree?)
- Bayesian inference: MrBayes for both genes
- Use model selection tools (e.g., ModelTest) to choose the best substitution model.
- Run ML trees with bootstrapping and Bayesian trees with posterior probability estimation.

Step 5: Compare the results

- Tree topology: Use Robinson-Foulds distance to measure differences.
- Bootstrap support: Compare average bootstrap percentages across methods.
- Clade consistency: Evaluate whether major clades appear consistently across trees.
- Rooting and resolution: Assess how well the tree separates lineages and whether it's well-resolved.
- Biological relevance: Compare results to known taxonomy or clinical patterns.

End!