**Understanding phylogenetic relationships between species of the Staphylococcus genus.**

**Introduction**

The genus *Staphylococcus* contains more than 60 species and subspecies. Many of these species lead to high levels of infection among human populations. Others are responsible for agricultural losses within the dairy, swine and poultry industries. Thus, the genus *Staphylococcus* is of interest to both the health and agricultural economic sectors. Since multiple species within this genus are common pathogens in non-human animals it becomes necessary that they be monitored with concern as these animals provide reservoirs for pathogenic bacteria. Host switching is an important mechanism in the evolution of *Staphylococcus*. For example, in *S. aureus*, human-to-poultry and bovine-to-human host switches have been observed. As such, a thorough understanding of species relatedness is a necessity for understanding host-pathogen interactions within this genus. The present study is aimed at determining evolutionary relationships between *Staphylococcus* species. Traditionally 16s rRNA sequences have been used for phylogenetic reconstruction but they have led to poor resolution of phylogenies based on this single gene dataset. This approach will include using a multi-locus dataset to get better to get better hierarchical and evolutionary relationships between species.

**Objective**

The broader objective of the study is to infer phylogenetic relationships between known species of the *Staphylococcus* genus using 16S rRNA gene sequence data. However, recent reports have revealed that 16S rRNA gene sequences alone are not sufficient to obtain a superior resolution of evolutionary relationships between different species. This study therefore also aims to analyze the role of multi-locus datasets (rpoB, tuf and dnaJ) to thoroughly explore the phylogenetic signal and provide robust confirmatory evidence for the relationships among *Staphylococcus* species. The analysis will involve multiple sequence alignments of the above-mentioned sequences followed by re-construction of individual phylogenetic trees using this data. This objective will be met by phylogenetic reconstruction using neighbor joining and maximum likelihood method using PAUP and Mr. Bayes/RaxML.

**Experimental plan**

DNA sequences for a total of four genes from 60 staphylococcal species, and two outgroup species (*Macrococcus caseolyticus* and *Bacillus subtilis*) will be downloaded from NCBI's GenBank. The four loci information will include the non-coding 16S rRNA gene sequences, and the three protein coding genes: dnaJ, rpoB, and tuf. Two types of trees – the neighbor joining and maximum likelihood trees will be created using PAUP and Mr.Bayes/RaxML. Neighbor joining applies a cluster algorithm to distance matrices to arrive at a fully resolved phylogeny. Maximum likelihood trees on the other hand use character-based methods to simultaneously compare all sequences in the alignment, considering one character (a site in the alignment) at a time to calculate a score for each tree.

**Expected Outcome**

Previously unreported relationships between Staphylococcus species may be an outcome of using a multi-locus dataset. The use of several phylogenetic methods will help infer a robust clustering of the different species in the phylogeny.

*Data Sources: NCBI GenBank*