Name(s)	Project Number
Dennis J. Shim	**S1523**

Project Title

Microsatellite Distribution Analysis in Various Bacterial Genomes Using Mathematica

Abstract

Objectives/Goals
This project analyzes various genomes to explore possible relationships between different bacteria based on microsatellite distribution. The objective of this research is to design and implement an algorithm into Mathematica to count the number of microsatellites in various genomes and subsequently examine distinctive patterns and characteristics found between and in bacterial families.

Methods/Materials
The genomes of different bacterial species were found in the NCBI database. An algorithm that could read these genomes in and count the number of microsatellites was designed and subsequently implemented into Mathematica. These results were plotted to find distinctive patterns, and the k-values, or slopes, were compared between species. The most frequently occurring microsatellites were also analyzed between species.

Results
It was found that there was an exponential correlation when the algorithm's output was plotted based upon rank of frequency versus the number of microsatellites detected. After linear regression was applied to the log graphs of the plots, the graphs of microsatellite length four showed the best correlation. In six species belonging to the family Enterobacteriaceae, the k-values fell within the range of 0.011 to 0.014, whereas the k-values for six species of the family Clostridiaceae were in the range of 0.025 to 0.030. It was also observed that many of the top ten most frequently observed microsatellite sequences were shared by various bacterial species of the same family.

Conclusions/Discussion
The data showed that the k-values of the species belonging to the same bacterial family stayed within a small range, and that range varied with the family. In addition, many of the most frequently observed microsatellites of a species were found to appear in many other genomes of species of the same family. Thus, the range of k-values and most frequently occurring microsatellites were found to be characteristics of species within a bacterial family. Therefore, microsatellite distribution can be used as a potential tool to differentiate bacterial species by family.

Summary Statement

This study identified two distinctive characteristics related to microsatellites that can be used as potential tools in differentiation of bacterial families.

Help Received

Ms. Lisa Fox helped review my paper; Dr. James Li advised my algorithm design and implementation.