CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)
Winnie C. Wang

Project Number
S2215

Project Title
Identification of Fly Groups through Microsatellite Frequency Analysis

Abstract

Objectives/Goals
Microsatellites are two to six base pair length sequences repeated up to hundreds of times within a genome. This project analyzes twelve different fly genomes to find a relationship between fly groups and their microsatellite distribution. The objective of the project is to create a Java program that counts the number of microsatellites and to use computational analysis to observe a pattern between the microsatellite distribution and the fly group.

Methods/Materials
Twelve fly genomes were obtained from the NCBI Flybase Database, and they were subsequently read into a Java program that counted the number of microsatellites and ranked each microsatellite by its prevalence in the genome. The output was then exported to Excel, where the results were graphed. The k-values were then compared among the fly groups for different microsatellite base lengths.

Results
The results reflected a negative exponential correlation between the microsatellite frequency and the fly group. The logarithmic frequency vs. rank was plotted, the k-value comparison indicated that microsatellite of length two was most accurate for group identification. The k-values of the melanogaster group ranged from -0.0798 to -0.0648, and the k-values of the obscura group ranged from -0.0431 to -0.0363.

Conclusions/Discussion
Because the k-values for fly species belonged to a certain range for each group, it is therefore possible to identify the group of an unknown fly with length two microsatellites. It can also be noted that microsatellites of length two are most frequent, and it may be possible that the microsatellite length for identification is species dependent. This test can be expanded to more advanced species, and if the correlation between microsatellite length and the group of species continues, then microsatellites can be a useful tool in species identification.

Summary Statement
This project demonstrated that microsatellite frequency of length two microsatellites can be used as a potential tool for identification of fly groups.

Help Received
Mrs. Nga Ngo was my advisor in this project, Mr. Joe Coglianese gave advice on Java data input, and Dr. James Li gave suggestions for methods of data analysis.