CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY

Name(s)

William Du

Project Number

S0509

Project Title

Novel Insights into Ultraconservation in Human Genes

Abstract

Objectives/Goals

In my research, my goal was to determine if there were nucleotide sequences that are 100% identical between the human cDNA and other organisms. If so, I would determine what these sequences are responsible for and attempt to find ultraconserved sequences not properly annotated in the human genome.

Methods/Materials

Highly conserved regions between animals were located through a series of BLAST searches. Regions were then selected if they had a Poisson value below 1*10^(-100). In one aspect of my research, DAVID, a gene annotation tool, was used to group the genes into clusters determined by function. In the second aspect, I searched for ultraconserved exonic regions that are in the predicted track of our genome but are not present in the known track. By using a second BLAST search, I discovered genes that are highly conserved between another organism's known track and the human#s predicted track and should be taken into consideration in the next revision of the human genome.

Results

Overall, over 10,000 sequences were identified to be ultraconserved. Subsequently, genes containing these sequences are predominantly involved in essential central dogmatic function. In addition, several genes upon closer inspection appear to play a key role in development and tumor suppression. Finally, two genome incongruences were discovered in the human genome; due to the correlation between the function of human genes and other organism genes, there is a high chance that this section of DNA should be annotated into the human known track.

Conclusions/Discussion

The human, chicken, mouse, frog, platypus, and sea squirt last shared a common ancestor hundreds of millions of years ago. Genetic drift causes the nucleotide sequences of genes to change even if their protein sequences remain essentially identical. Regions displaying 100% nucleotide identity over long (greater than 100 base pairs) stretches are statistically unlikely and may indicate the presence of previously unknown genomic functions. This research provides insight into highly conserved nucleotide sequences across various genomes in order to identify the function of these regions, mainly central dogmatic functions, and misannotations in the human genome. The results of the experiment are the foundation of research in potential gene therapy of diseases and provide understanding of genes that are ultraconserved.

Summary Statement

In my research, I discovered ultraconserved sequences between human cDNA and other organisms in order to link them to their function and better annotate the human genome.

Help Received

Used lab equipment at UC Davis Genome Center under supervision of Dr. Ian Korf, Participant in UCD Young Scholars Program