

## Public Abstract

First Name:S.

Middle Name:Jeffrey

Last Name:Reneker

Adviser's First Name:Chi-Ren

Adviser's Last Name:Shyu

Co-Adviser's First Name:Dmitry

Co-Adviser's Last Name:Korkin

Graduation Term:FS 2010

Department:Computer Science

Degree:PhD

Title:REFINED REPETITIVE SEARCHES AND LONG IDENTICAL MULTI-SPECIES ELEMENTS IN MAMMALS AND PLANTS - INSIGHTS INTO STRUCTURE, FUNCTION AND EVOLUTION

In this work, we have developed a tool to find exact matches to short segments of DNA whether it comes from mammals, plants, yeast or bacteria. Our searches are fast and accurate. However, since the information contained in DNA is written in a language that we do not understand, the results are of limited use by themselves. Therefore, our tool also searches the annotation data that accompanies the genomic DNA. In this way, life science researchers can quickly refine their single-species and/or cross-species searches. We used this approach to find something new in one species of bacteria based on what is already known from another. Also, some of the information within the DNA of related species is retained in exact copies over millions of years. We have found all of the exact matches in two sets of genomes. The first set is from human, mouse and rat and the other set is from Arabidopsis, soybean, rice, cottonwood, sorghum and grape. Some of our findings have never been seen before. For instance, we have identified the first sequences (> 200 base pairs) with exact copies in human, mouse and rat that are present multiple times in one or more of the three genomes. In our plant set, we found some new sequences that may be generated by specific enzymes in order to control multiple cellular functions. These findings are potentially valuable topics to continue to investigate.