

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

S. Jeffrey Reneker

Dr. Chi-Ren Shyu, Dissertation Supervisor
Dr. Dmitry Korin, Co-Advisor

ABSTRACT

All of the information necessary to reproduce a living organism is contained in the DNA of its genome. Within the genomic sequence, there are subsequences called genes that are transcribed into RNA and translated into proteins which control cellular function. The genomes of many different organisms, including human, have already been sequenced and many more are currently in progress. Also, there are concurrent projects to annotate the genes from these genomes. Once annotated, cross-species comparisons of both gene sequence and gene annotation terms are possible, which can facilitate knowledge discovery. This dissertation introduces our work in this area. For example, our newly-developed refined repetitive sequence searches were used to identify a potentially new phase-variable site in *Haemophilus influenzae* based on information previously reported from *Helicobacter pylori*.

The genome of an organism is inherited from its parents and is, in turn, passed on to its descendants. Determinative DNA sequences, both intragenic and intergenic, are often very highly conserved between diverging species over time. In fact, many of these sequences have been exactly conserved in multiple species throughout evolution. Once these sequences can be comprehensively identified in a set of genomes, they can be studied in more detail. This dissertation introduces our work in this area as well. For instance, several previously unreported long identical multi-species elements (LIMES) were identified and studied in mammals. Also, the comprehensive set of LIMES in six large plant genomes are identified and extensively characterized in terms structure, function and evolution.