LA-UR 11-06843

Licensable Technologies

SEQSTRAP Nucleic Acid Sequence Construction Software

Applications:

- Extrapolation from partially overlapping nucleic acid fragmented sequences to create full length sequences
- Biomarker identification
- Pathogen detection
- Robust assay design

Benefits:

- Preserves the biological knowledge and sequence variability available within fragmented samples
- More thorough analysis of nucleic acid sequences

Contact:

Ian Foti-Landis
505-665-6260
ifotilandis@lanl.gov

Kathleen McDonald
505-667-5844
kathleen_m@lanl.gov

tmt-4@lanl.gov

Technology Transfer Division

Summary:

For many diseases early detection is essential for improving a patient's treatment outcome. To support early detection, genomic sequence data can be "mined" to identify indicators (biomarkers) correlated to disease states. The key to designing a robust detection assay is the ability to search a collection of gene sequences to determine which sequences may provide a useful indicator for the target disease. However, much of the available gene sequence data is for fragments of the gene, which may generate misinformation, increasing the difficulty of identifying suitable and informative biomarkers. Furthermore, there are no software packages that can effectively utilize these fragments; therefore, these gene fragments are typically discarded, losing valuable information about a sequence's natural variability.

SEQSTRAP matches overlapping nucleic acid sequence fragments to extrapolate full length sequences.

SEQSTRAP was created to capture this information. SEQSTRAP reconstructs full length nucleic acid sequences from sequence fragments by using an iterative process to extend the sequence fragment into a full length sequence; utilizing the well understood occurrence of nucleotide base distributions. SEQSTRAP's sequence extrapolation process first compares all target sequences within the sample against each other, identifying the maximally similar pair with overhanging ends. The shorter sequence in the pair is extrapolated by adding the overhanging ends from its best matching partner. This process continues until no sequences further overlap and the target sequence stops growing, effectively recreating a full length nucleic acid sequence from fragments.

When SEQSTRAP is used with the co-developed software "PROSIG", it is possible to create, locate and identify specific nucleic acid sequences, allowing for the development of more robust diagnostic assays for difficult to detect pathogens.

SEQSTRAP is a parallel software written in C++ using MPI.

Development Stage:

A branched version of the software has been released as open source software. Development continues on an "in house" version.

Patent Status:

Software is copyright protected.

Licensing Status:

The software is available for exclusive or non-exclusive licensing.

Los Alamos NATIONAL LABORATORY ESTABLISHED 1943

www.lanl.gov/partnerships/license/technologies/

An Equal Opportunity Employer / Operated by Los Alamos National Security LLC for DOE/NNSA

11/11