

Splicing Graph Genome Assembler Software Modules

Tech ID: 19753 / UC Case 2002-822-0

TECHNOLOGY DESCRIPTION

This approach abandons the classical "overlap-layout-consensus" approach in favor of a new Eulerian splicing graph approach that, for the first time, resolves the problem of repeats in fragment assembly. The splicing graph approach, in contrast to the Celera assembler, does not mask repeats but uses them instead as a powerful fragment assembly tool.

UC San Diego is interested in commercializing its rights in the fragment assembly modules (see below). The research-quality software modules available are listed below. For general information about the EULER project, see [Pevzner, *et al*, PNAS, 98, 2001](#) and <http://nbcx.sdsc.edu/euler>.

- **EULER-Compare (SD2002-818)** consists of a Java user interface and a C server backend. It compares different sets of contigs, aligns them, and outputs information about the similarities between contig sets of different DNA sequence assemblies. The web-based Java user interface visualizes the comparison data as a contig-comparison graph.
- **EULER-Connect (SD2002-819)** is software that may be used to find some useful reads from the discarded reads to improve the assembly result and expedite the sequence finishing. EULER-connect can also verify chimeric reads.
- **EULER-EC (SD2002-820)** corrects errors in sequencing reads. For each read, it determines other overlapping reads and builds a multiple alignment. Using his multiple alignment, EULER-EC detects and corrects errors in the reads.
- **EULER-PCR (SD2002-821)** designs finishing multiplex PCR experiments for resolution of repeats that could not be resolved by sequence assemblers due to their length. Based on the repeat graph generated by EULER assembler ([Pevzner, *et al*, PNAS, 98, 2001](#)), the software identifies repeats and estimates their multiplicities. Every individual repeat is resolved by placing forward and reverse PCR primers at such distance from the beginning and the end of a repeat, so that all possible PCR products have different length. Thus, deducing the correct pairing between sequences outside of a repeat becomes a matter of measuring PCR product length by gel electrophoresis. EULER-PCR optimizes the number of reactions by pooling repeats that can be resolved simultaneously in a single multiplex PCR experiment.
- **EULER-TR (SD2002-822)** is software that improves the assembly result of EULER. Based on the repeat graph generated by the EULER assembler ([Pevzner, *et al*, PNAS, 98, 2001](#)), EULER-TR can resolve tangle edges (repeat edges) by inspecting the differences between reads fitted onto the tangled edge.

RELATED CASES

SD2002-822 includes technology cases SD2002-818, SD2002-819, SD2002-820, and SD2002-821.

CONTACT

University of California, San Diego
Office of Innovation and
Commercialization
innovation@ucsd.edu
tel: 858.534.5815.



OTHER INFORMATION

KEYWORDS

genome assembly, PCR, sequence
finishing, EST assembly

CATEGORIZED AS

- **Biotechnology**
 - Genomics
- **Computer**
 - Software
- **Research Tools**
 - Other

RELATED CASES

2002-822-0, 2002-818-1, 2002-819-1,
2002-820-1, 2002-821-1