

A Method and Software Significantly Improving the Accuracy Of Genome Assemblies: SEQuel

Tech ID: 22625 / UC Case 2013-803-0

BACKGROUND

Assemblies of next generation sequencing (NGS) data, while accurate, still contain a substantial number of errors that need to be corrected after the assembly process. Earlier assembly algorithms developed for Sanger sequencing follow an "overlap - layout - consensus" paradigm, where consensus refers to fixing errors in the contigs. Since this paradigm faces difficulties in short read assembly, most NGS assemblers employ a de Bruijn graph approach that effectively deals with large amounts of data. However, most NGS assemblers neglect the consensus step, *i.e.*, there exists no postprocessing of the contigs in Velvet and many other popular assemblers. Relying on high and uniform coverage, NGS assembly algorithms push the burden of producing high quality assemblies onto the construction of the de Bruijn graph. Our work demonstrates that NGS assemblers can benefit from the use of a consensus step. There are currently no tools that aim to accomplish this same goal.

TECHNOLOGY DESCRIPTION

UCSD researchers have recently developed a method and companion software, SEQuel, to correct errors (*i.e.*, insertions, deletions, and substitution errors) in the assembled contigs of NGS data. Fundamental of SEQuel is the *positional de Bruijn graph*, a graph structure that models *k*-mers within reads while incorporating the approximate positions of reads into the model. SEQuel takes as input an assembled contig, the paired-end reads that align to that contig and the approximate positions where they aligned, and returns a refined contig.

APPLICATIONS

Correcting errors in contigs from high throughput sequencing (HTS) assemblies. These might include bacterial/plant/vertebrate genomes that were not been previously sequenced, or the products of transcript assembly.

ADVANTAGES

- Removed 35% to 96% of small-scale assembly errors.
- Introduced positional de Bruijn graph for contig refinement.
- Demonstrated utility in hard (single-cell) assembly.
- SEQuel can be used in combination with any NGS assembler.

STATE OF DEVELOPMENT

The development stage of SEQuel is complete, and initial tests on several datasets have already shown its utility in correcting errors in assembled genomes. When SEQuel was applied to the recently assembled Deltaproteobacterium SAR324 genome, which is the first bacterial genome with a comprehensive single-cell genome assembly, and it made over 800 changes (insertions, deletions and substitutions) to refine this assembly.

RELATED MATERIALS

- Available upon request - Ronen R, Boucher C, Chitsaz H, Pevzner P. SEQuel: improving the accuracy of genome assemblies. *Bioinformatics*. 2012 Jun 15;28(12):i188-i196. - 06/15/2012

CONTACT

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



OTHER INFORMATION

KEYWORDS

next-generation sequencing, data assembly, SEQuel, post-processing of assembly, error correction, NGS

CATEGORIZED AS

- Research Tools
- Bioinformatics

RELATED CASES

2013-803-0

University of California, San Diego
Office of Innovation and Commercialization
9500 Gilman Drive, MC 0910, ,
La Jolla,CA 92093-0910

Tel: 858.534.5815
innovation@ucsd.edu
<https://innovation.ucsd.edu>
Fax: 858.534.7345

© 2012, The Regents of the
University of California
[Terms of use](#)
[Privacy Notice](#)