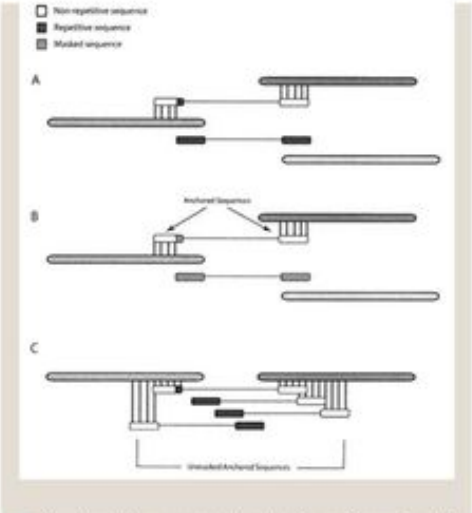
Technology Offer

KILAPE – Automated scaffolding and gap filling of large, complex genomes

Reference Number 11-00052

Challenge

The implementation of next-generation sequencing (NGS) technologies enabled massive throughput in the gathering of genomic information, coming along with the opportunity to use human genome data for medical and health care applications. However these new technologies produce reads that are shorter in length compared to traditional Sanger capillary-based electrophoresis systems. Especially highly repetitive sequences introduce limitations to the so far used de novo assembling methods, leading to poor coverage of repetitive regions and error prone results. Although de novo genome assemblers have seen extensive development to address this issue, stand-alone scaffolders and gap fillers using NGS data have remained largely uninvestigated.



(A) Unmasked sequences imply ambiguity in contigs pairing due to existence of repetitive sequences at the end of contigs.
(B) By masking repetitive sequences in the sequence reads, contig pairing due to repetitive sequences is eliminated and contigs pairing ambiguity is reduced. Masked reads that map to contigs are considered anchored.
(C) Unmasked anchored reads are clustered together with putative scaffolds and subjected to local assembly. Validating the putative scaffolds and performing implicit gap filling.

Technology

The KILAPE (K-masking and Iterave Local Assembly of Paired Ends) package provides an automated scaffolding and gap-filling software solution, improving the ambiguity arising from aligning repetitve segments. In short KILAPE provides tools for initial data processing, k-mer masking, read mapping to an initial *de novo* assembly, scaffolding, local assembly and concatenation. At this KILAPE predicts repetitive elements in NGS sequence libraries without resorting to a reference sequence. The algorithm produces fewer scaffolds with larger overall length and great long range validity. KILAPE can generally be used to improve assemblies produced by any *de novo* assembler, paving the way for higher quality genome drafts on lower-end hardware.

Commercial Opportunity

The technology is offered on www.genome.fli-leibniz.de/software/ as download under respective licensing conditions.

Leibniz-Institut für Alternsforschung – Fritz-Lipmann-Institut e.V.

Licensing Contact:
Dr Sabina Heim
Senior Technology Manager
T: +49 531 618120-90
heim@ascenion.de

Ascenion GmbH
Herzogstraße 64
D-80803 München
T: +49 89 318814-0
info@ascenion.de
www.ascenion.de