Paper Impact: Assembly of long, error-prone reads using repeat graphs

Johannes Hausmann[1], Luis Kress[1]

[1] Goethe Universität Frankfurt

Key References

- Assembly of long error-prone reads using de Bruijn graphs.
  - Authors: Yu Lin (First), Pavel A. Pevzner (Last and corresponding)
  - Journal: PNAS (113 (52), 8396-8405)
  - Year: 2016
  - Citations: 126 (Google scholar)
  - Reason: The presented algorithm Flye is an improvement of the tool presented in this article.
- De novo repeat classification and fragment assembly.
  - Authors: Paul A. Pevzner (First), Glenn Tesler (Last and corresponding)
  - Journal: CSH Press, Genome Research (volume 14, 1786-1796)
  - Year: 2004
  - Citations: 294 (Google scholar)
  - Reason: Basics on the repeat graph and the repeat classification problem.
- Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution.
  - Authors: Zhaoshi Jiang (First), Evan E Eichler (Last and corresponding)
  - Journal: Nature Genetics (volume 39, 1361-1368)
  - Year: 2007
  - Citations: 192 (Google scholar)
  - Reason: The information, that repeat graphs can be used to represent mosaic structures.
- What is the difference between the breakpoint graph and the de Bruijn graph?
  - Authors: Yu Lin (First), Pavel A Pevzner (Last and corresponding)
  - Journal: BMC Genomics volume 15, Suppl 6 (S6)
  - Year: 2014
  - Citations: 23 (Google scholar)
  - Reason: Groundwork publication from Pevzner. Information on the graph structure used to implement Flye. Assembly graphs created from the repeat graph are special cases of breakpoint graphs.
- Haplotype and Repeat Separation in Long Reads.
  - Authors: Tischler-Höhle G.
  - Journal: Computational Intelligence Methods for Bioinformatics and Biostatistics, CIBB 2017, 103-114
  - Year: 2017
  - Citations: 3 (Google scholar)
  - Reason: The publication describes methods for repeat & haplotype seperation with long reads. Resolving unbriged and highly similar repeats is realted to the challenge of overlap-filtering repeat resolution described in this publication.

Citations of our article

- Citations: 564 (Google scholar); 264 (PubMed); 273 (Web of Science); 356 (Cross Ref)
- Yearly
  - PubMed: 2019: 24; 2020: 141; 2021: 108
  - Altmetric: 2017: 1; 2018: 3; 2019: 58; 2020: 280; 2021: 199

Citations per year (Altmetric)

- Articles which cite the given paper:
  - Fast and accurate long-read assembly with wtdbg2 (Nature Methods) (cited by: 313, Google scholar), (impact score: 10.93, 2020)
  - De novo assembly of haplotype-resolved genomes with trio binning (Nature Biotechnology) (cited by: 151, Google scholar), (impact score: 10.71, 2018)
  - Telomere-to-telomere assembly of a complete human X chromosome (Nature) (cited by: 180, Google scholar) (impact score: 42.778, 2019)
  - Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome (Science) (cited by: 160, Google scholar), (impact score: 12.84, 2019)
  - Opportunities and challenges in long-read sequencing data analysis (Genome Biology) (cited by: 180, Google scholar) (impact score: 11.71, 2020)

Corresponding author Pevzner

- SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing
  - Reason: last author, 12583 citations, widely used assembly tool, (impact score 2.09; 2013)
- Initial sequencing and comparative analysis of the mouse genome
  - Reason: first sequencing and assembly of the mouse genome, 7438 citations, (impact score 36; 2010)
- De novo identification of repeat families in large genomes
  - Reason: corresponding author, groundwork for the given paper (Assembly of long-error prone reads) /groundwork for assembly of repeat rich genomes, 1211 citations, (impact score 5.15; 2013)
- An Eulerian path approach to DNA fragment assembly
  - Reason: first author, groundwork for further short read assembly tools, 1571 citations, (impact score 9.81; 2013)
- De Novo Peptide Sequencing via Tandem Mass Spectrometry
  - Reason: last author, development of de novo peptide sequencing which is still used today, 747 citations, (impact score 2.09; 2013)