Yilei Fu

Computer Science Ph.D. Student at Rice University, focusing on bioinformatics.

6100 Main Street, Houston, TX | yf20@rice.edu | Google Scholar | GitHub | Twitter

Education

Ph.D. Candidate| 2019.8 - PRESENT | Rice University

- Research at Treangen Lab, focusing on Structural Variants detection, long-read alignment and metagenomics.

B.E. In Computer Science | 2014.09.01-2018.06.31 | Harbin Institute of Technology (HIT), China

- Major: Computer Science
- Conducted research on Structural Variants detection in human genomes and read alignment.
- Major Weighted Average Score: 90.41/100 (4.0 in the scale of 4.0)

Exchange Student| 2017.1 - 2017.6| UC Berkeley

- Took courses on computer science area.

Skills & Abilities

Programming Language:

- Proficiency in Python and C++.
- Sufficiency in Java, R and VHDL.

Algorithmic Expertise:

- Proficiency in read/sequence alignment algorithms, Structural Variant detection algorithms and data compression algorithms.
- Sufficient in genome assembly, parallel computing and SNP calling algorithms.

Experience

Teaching and Research Assistant| 2019.8 - PRESENT | Rice University

- 2-year TA experience on a graduate level course “Introduction to computer graphics”.
- Research assistant at Treangen Lab. Conducted research related to SARS-COV-2.

Research Scientist | 2018.1 - 2019.6 | The Center of Bioinformatics at HIT

- Conducted research on Structural Variants detection in human genomes and read alignment.

Data Analyst| 2015.1| Institute of Extreme Environmental Nutrition and Protection at HIT

- Statistical analysis of experimental data of the paper Acanthopanax senticosus reduces brain injury in mice exposed to low linear energy transfer radiation.

Publications and Projects

Fu, Yilei, et al. "Vulcan: Improved long-read mapping and structural variant calling via dual-mode alignment." bioRxiv (2021).

- Improves structural variant calling by utilizing different gap penalty functions

SUEDE: Strain-level UniquE DNA Extractor (2019)

- https://gitlab.com/treangenlab/probefinder
- By finding maximum unique matches with positive DNA sequences and concatenated negative similar DNA sequences, we can extract unique DNA marker in the positive sequences, distinguishing them from those negative sequences.

Fu, Yilei, et al. "rCANID: read Clustering and Assembly-based Novel Insertion Detection tool." 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018.

- We used read clustering and genome assembly methods for detection of novel insertion in Human genome.

Zhan, Q., Fu, Y., Jiang, Q., Liu, B., Peng, J., & Wang, Y. (2020). SpliVert: a protein multiple sequence alignment refinement method based on splitting-splicing vertically. Protein and Peptide Letters, 27(4), 295-302.

- By splitting protein sequence into low quality – high quality – low quality regions, we improved alignment quality of protein multiple sequence alignments.

Guo, H., Fu, Y., Gao, Y., Li, J., Wang, Y., & Liu, B. (2019). deGSM: memory scalable construction of large scale de Bruijn Graph. IEEE/ACM transactions on computational biology and bioinformatics.

- We applied BWT on unitigs of *de bruijn* graphs to compress the memory requirement for their constructions.

Guo, H., Liu, B., Guan, D., Fu, Y., & Wang, Y. (2018, December). Fast variation-aware read alignment with deBGA-VARA. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 227-233). IEEE.

- We applied variation aware algorithm on our previous read aligner deBGA.

Zhou, A. Y., Song, B. W., Fu, C. Y., Baranenko, D. D., Wang, E. J., Li, F. Y., & Lu, G. W. (2018). Acanthopanax senticosus reduces brain injury in mice exposed to low linear energy transfer radiation. Biomedicine & Pharmacotherapy, 99, 781-790.

- I have performed statistical data analysis in this paper.

Awards

Ken Kennedy Institute Computer Science Engineering Enhancement Fellowship, 2019