Eliminating Heterozygosity from Reads Through Coverage Normalization

**ABSTRACT**

Heterozygosity has long plagued genome assembly. A wealth of sophisticated algorithms and procedures have been introduced for their treatment in genome assembly. In this paper, we propose a method called Hank (Heterozygosity assimilation through normalized k-mers) for this purpose. The method eliminates heterozygosity from reads by modifying the k-mers in heterozygous regions to increase their coverages to the levels of those in homozygous regions. When evaluated on simulated Illumina data at levels of heterozygosity from 0.1% to 2.0%, Hank was able to remove 80–96% of the heterozygosity. We also examined the effects of the corrections on de novo genome assembly using SOAPdenovo2, ALLPATH-LG and Platanus. All three methods improved in performance using the treated k-mers (we do not include these assembly results in this manuscript due to space constraint).

This paper was presented at The IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Dec 15-18, 2016, Shenzhen, China.

Supervisor: Dr Shuaicheng LI
Research Interest: Bioinformatics

All are welcome!

In case of questions, please contact Dr Li Shuaicheng at Tel: 3442 9412, E-mail: shuaicli@cityu.edu.hk, or visit the CS Departmental Seminar Web at http://www.cs.cityu.edu.hk/news/seminars/seminars.html.