JUSTIN R. FISHER, Dept of Computer Science, Mathematics, and Engineering, Shepherd University, Shepherdstown, WV, 25443, and JASON R. MILLER, Dept of Computer Science, Mathematics, and Engineering, Shepherd University, Shepherdstown, WV, 25443. Efficient Simulation of Allele-Specific Expression.

Diploid organisms such as animals and plants carry maternal and paternal variants of most of their genes. Preferential transcription of either gene variant is called ASE for allele-specific expression. In plant seeds, ASE has been observed at selective genes at selective developmental stages, so the process is presumably regulated by epigenetic factors such as genomic imprinting. The Informative Reads Pipeline (IRP) is software that we developed previously for the purpose of detecting ASE in RNA sequencing data obtained from plant seeds. To help us validate and generalize the software, we developed a sequence data simulator that harbors a parameterized model of ASE. Whereas the maternal/paternal ratio per gene is always unknown in real data, the simulator provides the opportunity to quantify IRP's ability to recover the preset ratios from the data provided. The simulator generates and maps sequences using standard software. Simulating ASE at all combinations of all genes would be computationally prohibitive. Therefore, we introduced an optimization that reduces the generate+map computation from exponential to constant time. Correctness of the optimized simulator is demonstrated here.