

## TURNING REPEATS TO ADVANTAGE: SCAFFOLDING GENOMIC CONTIGS USING LTR RETROTRANSPOSONS

A. Kalyanaraman, S. Aluru and P.S. Schnable  
Department of Electrical Engineering and Computer Science  
Departments of Agronomy, and Genetics, Development and Cell Biology  
Iowa State University,  
Ames, IA 50011, USA

The abundance of repeat elements in the maize genome complicates its assembly. Retrotransposons alone are estimated to constitute at least 50% of the genome. In this paper, we introduce a problem called *retroscaffolding*, which is a new variant of the well known problem of *scaffolding* that orders and orients a set of assembled contigs in a genome assembly project. The key feature of this new formulation is that it takes advantage of the structural characteristics and abundance of a particular type of retrotransposons called the *Long Terminal Repeat (LTR) retrotransposons*. This approach is not meant to supplant but rather to complement other scaffolding approaches. The advantages of retroscaffolding are two fold: (i) it allows detection of regions containing LTR retrotransposons within the unfinished portions of a genome and can threefold guide the process of finishing, and (ii) it provides a mechanism to lower sequencing coverage without impacting the quality of the final assembled genic portions. Sequencing and finishing costs dominate the expenditures in whole genome projects, and it is often desired in the interest of saving cost to reduce such efforts spent on repetitive regions of a genome. The retroscaffolding technique provides a viable mechanism to this effect. Results of preliminary studies on maize genomic data validate the utility of our approach. We also report on the on-going development of an algorithmic framework to perform retroscaffolding.