

Abstract ID:- 186

Abstract Topic:- Genetic, genome and epigenome databases and resources

Abstract Title:- T-ReX:Tandem Repeat identification using XOR operations

Presenting author name :- Anukrati Sharma

Presenting author institute:- CSIR-Centre for Cellular & Molecular Biology, Hyderabad

Co-authors name:- Akshay Avvaru, Divya Tej Sowpati

Co-authors institute:- CSIR - Centre for Cellular and Molecular Biology, Hyderabad; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad

Aims:-Tandem repeats (TRs) are widely prevalent in eukaryotic genomes, and account for 3% of the human genome. They are linked to varied biological functions and their unique length polymorphism contributes to phenotypic variation both within and among species. Although TRs are well studied in terms of the motif length, motif composition and genomic context, the relevance of purity remains less understood. A possible bottleneck for this is the lack of an efficient tool that can identify imperfect tandem repeats. We set out to address this limitation with a fast and accurate method for identifying imperfect TRs.

Methods:- The basic concept behind identifying repeat regions involves a process of shifting and matching the DNA sequence with itself. Using binary representation of DNA sequences as dynamic bitsets, we perform the shift-and-match step for the entire chromosome in one go. TR identification begins by locating “seeds” that exhibit high bit counts using a sliding window approach. Each seed is then aligned with a pseudo-perfect repeat of the same length using a highly optimized SIMD execution of the Smith-Waterman algorithm. Those alignments which match the user-defined criteria are reported in the output file.

Results:- Our tool “T-ReX” focuses on faster and accurate identification of imperfect TRs that doesn't require advanced computational power. In a comparative analysis against other tools like Dot2Dot and TRF, T-ReX identified more TRs, including both perfect and imperfect repeats, and was faster even when run on a standard CPU with minimal RAM requirements. T-ReX offers accuracy and rapid performance, along with flexibility by allowing user defined parameters.

Conclusions:- T-ReX provides efficient and rapid processing of sequences as it utilizes binary representations of bases operating through dynamic bitsets and SIMD execution of alignment to parallelize the process. T-ReX can detect and report TR sequence mutations with tailored configurations that allows users to customize their research.

Keywords:- Tandem Repeats, Algorithm, Mutations