

## SEEDER: PERL MODULES FOR CIS-REGULATORY MOTIF DISCOVERY

François Fauteux<sup>1,2</sup>, Mathieu Blanchette<sup>2,3</sup> and Martina V. Strömvik<sup>1,2</sup>

<sup>1</sup>Department of Plant Science, McGill University, Sainte-Anne-de-Bellevue, Canada, H9X 3V9

<sup>2</sup>McGill Centre for Bioinformatics, McGill University, Montreal, Canada, H3A 2B4

<sup>3</sup>School of Computer Science, McGill University, Montreal, Canada, H3A 2A7

The computational identification of *cis*-regulatory elements in DNA sequences is a notoriously difficult problem in contemporary bioinformatics. *Cis*-regulatory motifs are typically short (6-15 bp), somewhat degenerate and often hard to distinguish from the background sequences.

We have recently released the Seeder 0.01 suite of Perl modules implementing an exact, discriminative seeding algorithm designed for fast and reliable DNA motif discovery in the promoter sequences of co-regulated, functionally related or homologous eukaryotic genes. The algorithm outperforms popular motif discovery tools on biological benchmark data.

The Seeder algorithm uses an enumerative approach and an objective function based on the probability of the sum of Hamming Distances between words (combinations of nucleotide symbols) and best matching subsequences (substring minimal distance, SMD), given a word-specific background probability distribution. This computation is accelerated by using the SMD index, a data structure allowing an efficient lookup, in a given sequence, for a subsequence minimally distant to a given word.

We present an application of the algorithm to the identification of plant-family-specific, tissue-specific *cis*-regulatory promoter motifs. DNA motifs discovered in the promoter sequence of tissue-specific genes are highly similar to experimentally characterized plant *cis*-regulatory elements.

We also present a web server for motif discovery using the Seeder algorithm, with pre-computed background for six plant species. Various tools are also accessible on the Seeder website for post-processing operations including imaging, PWM scoring and database matching.

Reference: Fauteux F, Blanchette M, Stromvik MV: Seeder: Discriminative Seeding DNA Motif Discovery. *Bioinformatics* 2008, 24(20):2303-2307.

Seeder website: <http://seeder.agrenv.mcgill.ca>

Seeder 0.01 distribution: <http://search.cpan.org/~ffauteux/Seeder-0.01/> (released under the terms of the Artistic License)

E-mail address of the presenting author: [francois.fauteux2@mail.mcgill.ca](mailto:francois.fauteux2@mail.mcgill.ca)