Volume 50(3-4):145, 2006
Acta Biologica Szegediensis
http://www.sci.u-szeged.hu/ABS

DISSERTATION SUMMARY

Usage of enumeration method based algorithms for finding promoter motifs in plant genomes

Mátyás Cserháti

Institute of Plant Biology, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

Understanding the underlying genetic and physiological processes would make it possible to enhance crop yield in the face of abiotic stress. To do so would entail unraveling the complex regulatory interactions behind this phenomenon. The analysis of transcription factor binding sites in promoters is a common and useful tool in such studies.

Our work focused on developing a number of computer algorithms for finding motif dyads in a set of input promoter sequences for co-regulated genes, since they tend to group together into regulatory complexes in promoters. Our algorithms belong to the family of algorithms called enumeration methods, which perform an exhaustive search for all possible motifs in the input sequences. In our case we looked for dyad motifs, described by the formula $M_HN_nM_T$, where M_H and M_T are individual motifs, each the same length, and N_n represents a spacer between the head and tail motifs n bp long.

For each algorithm we counted the number of head and tail motifs as a function of spacer length. The basic logic behind each algorithm was to quantify the difference between the distance distribution of a given dyad and the uniform distribution. We assumed that if a given dyad is biologically irrelevant, then its head and tail motifs would occur randomly at all distances from each other in equal numbers, thereby producing a nearly uniform distribution. If the dyad has a biological function, then they should occur at a given distance from each other in high numbers. In this case, the distribution of the head and tail motifs should depart from the uniform distribution, and accumulate at a specific distance from each other.

The first and second algorithm measures the difference between the maximal occurrence and the average occurrence of the dyad according to motif distance. Greater differences infer biological relevance (Cserháti 2005). The third algorithm represents individual motif distances with “boxes” into which a number of “balls” are distributed, which represent the occurrences of the dyad at that motif distance. The algorithm calculates the probability of this distribution. Less probable distributions can be assumed to be more relevant. The fourth algorithm calculates the level of homogeneity between the distribution of the dyad in the input promoter set and a set of randomly selected promoters (Cserháti 2006).

We tested our four algorithms on 130 *Arabidopsis* stress genes, and searched for pentamer dyads with a maximum distance of 52 bp in between. The top 50 dyads were selected for each method, and the *Arabidopsis* promoterome was screened to find promoters with high motif content. 72.3% of the original promoters could be found back with our methods, and 58.7% were found to be involved in stress. The fourth method has already been experimentally verified by comparing the number of dyads in the family of rice aldo-keto reductase genes.

Another work involved reviewing 151 CDK genes in plants. Overall, 29 genes were found in *Arabidopsis*, and 30 in rice. Promoter analysis was performed for 26 of these genes. More motifs were found in rice, which were involved in ABA and auxin response. Ethylene response elements were also found in CDKB’s and CDKC’s, but not in CDKA’s. A number of elements were found which are involved in light response and Circadian rhythm. The MSA, MYB, and APETALA3/AGAMOUS motifs were also found in a large number of promoters.

References

Cserháti et al. (2005) Statistical methods for finding biologically relevant motifs in promoter regions and a few of its implementations. Proceedings of the 5th International Conference of PhD Students. Miskolc. ISBN 963 661 681 6:41-46.

Cserháti et al. (2006) Enumerációs módszereken alapuló algoritmusok használata promóter motívumok keresésére. Tavaszi Szél 2006 conference. Kaposvár. ISBN 963 229 773 3.

Supervisor: János Györgyey
E-mail: csmatyi@nucleus.szbk.u-szeged.hu