Sequence analysis

FineStr: a web server for single-base-resolution nucleosome positioning

Idan Gabdank1−*, Danny Barash1 and Edward N. Trifonov2,3

1Department of Computer Science, Ben Gurion University of the Negev, Beer Sheva 84105, 2Genome Diversity Center, Institute of Evolution, University of Haifa, Haifa 31905, Israel and 3Division of Functional Genomics and Proteomics, Masaryk University, Kamenice 5, Brno CZ-62500, Czech Republic

Associate Editor: Limsoon Wong

ABSTRACT

Summary: The DNA in eukaryotic cells is packed into the chromatin which is composed of nucleosomes. Positioning of the nucleosome core particles on the sequence is a problem of great interest because of the role nucleosomes play in different cellular processes including gene regulation.

Using the sequence structure of 10.4 base DNA repeat presented in our previous works and nucleosome core DNA sequences database, we have derived the complete nucleosome DNA bendability matrix of Caenorhabditis elegans.

We have developed a web server named FineStr that allows users to upload genomic sequences in FASTA format and to perform a single-base-resolution nucleosome mapping on them.

Availability: FineStr server is freely available for use on the web at http://www.cs.bgu.ac.il/~nucleom. The site contains a help file with explanation regarding the exact usage.

Contact: gabdank@cs.bgu.ac.il

Received on November 17, 2009; revised on January 3, 2010; accepted on January 20, 2010

1 INTRODUCTION

DNA accessibility plays an important role in the regulation of processes such as transcription, replication, recombination and DNA repair. DNA accessibility in turn depends on the chromatin structure, of which the building blocks are nucleosomes.

It is generally accepted that the segment of DNA double helix which is wrapped around the histone octamer forming the nucleosome should possess sequence-dependent bendability properties that would provide a stabilizing contribution to the free energy of the nucleosome 3D structure.

We have discovered the sequence pattern characteristic for nucleosome DNA of Caenorhabditis elegans and, apparently, applicable to the nucleosomes of all eukaryotes (Trifonov, 2009; Trifonov, 2010). As described in the cited work, the C.elegans pattern is identical to the pattern uniquely defined by DNA deformability properties. In particular, it is dictated by minimization of unstacking of DNA bases and base pairs in the nucleosome DNA. Accordingly, we built a web server FineStr [pronounced as (fi-nes’tru), from Latin fenestra—window, and from fine structure] that users worldwide could use in order to map nucleosomes to genomic sequences of interest with a single-base-resolution.

2 METHODOLOGY

Crystallography data suggest that nucleosome DNA sequence is wrapped around histone octamer, having 12 close contacts of oriented inward minor groove with arginines of the histones (Arents et al., 1993; Harp et al., 2000). That makes at least 11 periods of DNA helix turns tightly wrapped around the octamer. To reveal the full 11 periods bendability matrix, we used the previously published 10.4 base nucleosome DNA repeat sequence structure (Gabdank et al., 2009) and nucleosome core DNA sequences database (Johnson et al., 2006).

We concatenated the 10-matrix 11 times, placing the nucleosome dyad axis on the base rather than on interbase (Davey et al., 2002). In order to make it we had to duplicate the central matrix element, dinucleotide AT, and to place it in the positions −1 and 0, by that we ensured preferential positioning of the dinucleotide in the positions −1, 0 and 0, 1. Similarly the local dyads at the positions 26 and −26 we have duplicated the matrix element CG and at the positions 52 and −52 the AT dinucleotide. The base central local dyad positions 26 and 52 correspond to 2.5 and, respectively, five 10.4-base period distances from the main dyad of the nucleosome DNA.

To accomodate the stretching/compression of nucleosome DNA by 1 to 2 bases (Ong et al., 2007), we have derived two additional matrices M−1 and M+1, by shifting the original matrix one base left and right. These matrices were combined to produce the Mcomb matrix by choosing the maximal value for each position Mcomb(i) = max{M−1(i), M0(i), M+1(i)}.

The nucleosome mapping probe Mfinal was completed by symmetrization of the combined matrix Mcomb. Values for complementary dinucleotides (AAC/TC, AGGC and CTTA) were taken from the i-th (for any given dinucleotide XY) and (116−i)-th positions (for its complementary dinucleotide), averaged and put back in the i-th and (116−i)-th positions of the Mfinal. For the self-complementary AT and CG dinucleotides, the symmetrized values were calculated by averaging the values of the i-th and (116−i)-th positions for the given dinucleotide.

The suitability of any DNA sequence to the binding in the nucleosome structure can be evaluated using the probe Mfinal simply by aligning the matrix with the sequence and summing the scores at all 116 matrix positions. The ‘strength’ of the nucleosome that would be formed by wrapping DNA with the given sequence on histone
All numerical calculations are performed with PHP code that uses the complete bendability matrix data to perform the nucleosome mapping on a given sequence. The presentation interface was implemented using HTML and JavaScript on the client side and PHP on the server side. In order to generate charts, we have used the GFlot package.

Computational time (seconds) depends on the sequence length presented to the server, and this length is currently limited to 10,000 letters to reduce the informational load on the clients’ browser that is responsible for the presentation of the charts of the nucleosome positioning predictions using javascript.

The FineStr server web interface enables the user to submit the sequences of interest for analysis either by pasting the sequences into the text box or by uploading a file containing the sequences in FASTA format. For each sequence submitted, the server generates the nucleosome probe score distribution along the sequence, pointing to the places with higher likelihood to place a nucleosome. The server response includes a graphical representation of the results (Fig. 1) and an option to download a file containing the numerical representation of the results. The plot in Figure 1 displays a typical example of two nucleosomes mapped on sequence chosen for the illustration purpose from the first chromosome of Caenorhabditis elegans. Alternative nucleosome center positions are seen around the central maxima. This corresponds to physical reality—sliding of the nucleosomes to alternative positions with the same rotational setting (Bao et al., 2006).

5 RESULTS

Crystallography at atomic resolution at present time is the most accurate way to map nucleosome position on the sequence. Respectively, any other mapping technique to be tested should use the crystallography data as reference. In order to test our sequence-directed algorithm that uses the nucleosome mapping probe, we took all known nucleosomes crystalized with unique sequences, where the central base is determined with atomic resolution. These are protein data bank (PDB) structures: 1AOI (Luger et al., 1997), 1EQQ (Harp et al., 2000), 1KX4, 1KX5 (Davey et al., 2002) and 2NZD (Ong et al., 2007). Comparison of the nucleosome positions calculated by our probe with the crystallography data showed an exact match within ±1 base.

6 FUTURE DEVELOPMENTS

The software allows for modifications and extensions, e.g. expected small species specific changes in the matrices. Another interesting extension would be the incorporation of modified matrices that can help in the detection of various functionally active special nucleosomes positions (e.g. epigenetic nucleosomes). ACKNOWLEDGEMENTS

The authors would like to thank Yuval Farkash, Orit Wald and Ori Klein for providing technical support with setting up the server. Funding: Pratt fellowship; Lynn and William Frankel Center for Computer Sciences at Ben Gurion University. Conflict of Interest: none declared.

REFERENCES