Sequence Logo Visualization based on Gestalt Perception – Novices vs Experts

Weiying KOK[a], Yin Bee OON[b] and [*]Nung Kion LEE[c]

[a,b,c] *Faculty of Cognitive Sciences and Human Development*
Universiti Malaysia Sarawak
{yinbee[b], [*]nklee[c]}@fcs.unimas.my

ABSTRACT

Sequence logo is a popular graphical method for displaying the conservation characteristics of a sequence motif profile. Studies have found that decision making biases and misused of sequence logo occur due to the differences in individual perception, knowledge and experiences. The aim of this study is to identify the differences between the novices and experts in visualizing sequence logo based on Gestalt perception. 52 participants were involved in an online survey and independent t-test was used to analyse the data obtained. Results shown that there are significant differences in the perception and needs between novice and expert users. Visual cues and detailed information display are required by novice users whereas experts prefer a simple but more functional representation in the sequence logo. Therefore, it is essential to improve the sequence logo visualization in the colour, grouping, information display and interactivity of the tool in order to support the needs of various users.

KEYWORDS

Sequence Logo; sequence motif; novice; expert; Gestalt perception;

INTRODUCTION

Nowadays, tools for discovering and visualizing sequence motifs have become essential for life scientists in solving various motif finding problems. Biological sequence motifs are short recurring sequence patterns that represent many features in DNA, RNA and proteins. Discovering and visualizing sequence motifs are very important in biological field as these lead to a better understanding of the transcription regulation, splicing of mRNA and the formation of protein complexes (Bailey, 2008). Visualization tools are important in helping researchers to visualize results in order to attain better scientific insight or to infer significant conclusions. A simple visualization tool is capable to provide powerful, self-explanatory messages in the data analysis. In 1990, Schneider and Stephens presented sequence logo to help user in the interpretation of the sequence data by visualizing the conserved motif that represents various structural or functional properties.
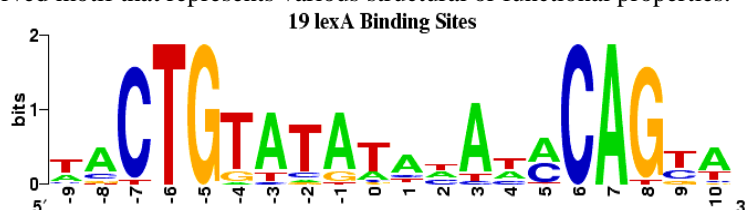

Figure 1 An example of E. coli transcription factor binding sites sequence logo generated by Weblogo

In sequence logo, individual nucleotide or amino acid characters representing the sequence are stacked on top of each other by positioning the most frequent nucleotide on top of the sequence stack (Schneider & Stephens, 1990). The height of each nucleotide symbol in the stack is proportional to its frequency and is determined by the total information content (measured in bits) at each position in a sequence. The information presented at each position of the multiple sequence alignments are calculated based on Shannon's information theory (Schneider & Stephens, 1990). A sequence logo can represent the information on: (a) the general consensus of the sequences; (b) the predominance order of the nucleotide / amino acid at each position; (c) the relative frequency of each nucleotide / amino acid; (d) the amount of information present at each position in the sequence; and (e) the initiation point or cut point of the sequences; into a single graphical representation (Schneider & Stephens, 1990). Understanding visual perception can be rather subjective because one with different level of education, culture, experience or exposure would perceive them differently. The choice of visual attributes, visual cues and presentation of information will also affect the way people perceive the information represented. Hence, it is essential to design visualization with the understanding of visual perception in relation to size, colours, orientation and etc. to help to improve the quality and quantity of information displayed.

_____

* Corresponding author