Perceptual Principles for Scalable Sequence Alignment Visualization

Danielle Albers*
University of Wisconsin - Madison

Michael Gleicher†
University of Wisconsin - Madison

1 Introduction

Sequence comparison is a fundamental task in the biological sciences. Scientists often need to understand the similarities and differences between genetic sequences to understand evolution, to infer common function, or identify differences. Because the sequences are too long for manual examination, scientists rely on alignment tools that automatically identify subsequences that match between the sequences being compared. Numerous approaches for displaying and exploring alignments exist, and have been incorporated into a wide variety of tools. See [Procter et al. 2010] for a survey of several existing approaches.

The quantity and complexity of genomics data is growing rapidly, increasing the challenges for data interpretation tools. Most obviously, tools must handle longer genomes and comparison between more genomes. Alignments might contain many-to-many correspondences, probabilistic information, and other types of complex data. Developing methods for presenting this data is difficult: systems will need to present vast amounts of information to the user. Some scalability limits are explicit in system designs or come from memory or performance issues. But often scalability is hindered because the visual design breaks down.

Our conversations with users suggest that existing tools break down as the interpretation problems scale up: either the displays become too complex or too cluttered to be interpreted effectively. This is not surprising as the human visual system is limited in the amount of information it can take in. By understanding these limits, we hope to design systems that can work within them.

Our goal is to build new visualization tools that will better scale to the challenges of modern genomics. As a first step in this direction, we have been studying how results from perceptual science can inform the design of alignment visualization tools. This poster surveys how perceptual principles are relevant in current alignment visualization approaches, either in understanding scalability limitations or in suggesting ways to improve the current tools. Ultimately, we feel that new designs based on these principles will offer the best strategy towards developing future alignment visualization tools that will be effective as the problem scales.

2 Perceptual Principles

Perceptual science has explored the capabilities and limitations of the human visual system. Here, we survey some results that seem relevant to the design of alignment visualization. While some categories of ideas are well-known to the visualization community, others are either more recent developments or particularly relevant to large-scale alignment visualization.

Pre-attentive phenomena allow a viewer to very rapidly identify targets in cluttered environments. Manipulating pre-attentive features within an image simplifies visual search by making certain groups of objects "pop out." When the system knows what the user is trying to find, it can use pre-attentive cues to highlight the targets. However, care must be taken to use highlighting effectively and to avoid unwanted pop-out.

*e-mail: dalbers@cs.wisc.edu
†e-mail: gleicher@cs.wisc.edu

Visual search occurs when a viewer cannot find targets pre-attentively and must scan their attention over the scene to search for targets. Without perceptual aid, search tasks can be cognitively heavy and time-consuming [Alvarez et al. 2007]. Task-specific search aids facilitate the perception of wider variations of data association. By designing aids which cooperate with perceptual mechanisms, users are better able to process the data for more rapid and efficient visual search.

Pre-search processing initiates the visual search process by collecting information to guide search. During the initial processing of a scene, pre-search processes gather structural and feature-based "proto-object" information to develop a contextual map of a scene. Blending this scaffolding with the relationships amongst the proto-object data results in more rapid identification of regions of interest and more efficient searches over the visual space.

Visual clutter is when item quantity, encoding, or layout causes performance degradation for search tasks. Visual clutter impairs the perceptual system by bogging down cognitive processes and slowing visual search. In data-processing tasks like sequence alignment comparison, clutter reduction by adjusting semantic data granularity often proves far more effective than simply removing data, while still preserving the overall data set [Rosenholtz et al. 2007].

Summarization is the ability to construct statistical summaries of non-attended regions [Balas et al. 2009]. Summarization phenomena offer an opportunity to rapidly provide overview information without requiring the viewer's specific attention, which should be valuable in alignment visualization as often a scientist needs the context, not the details, of objects outside their immediate focus.

While many of these principles can and have been directly applied to improve current designs, we feel that together they suggest a new generation of tools that use multiple, coordinated views each designed to exploit pre-attention, simplify visual search tasks, decrease visual clutter, and allow for rapid visual summaries. In the future, we hope to design and build such a system.

Acknowledgements: Funding for this project was provided by the US Department of Energy Genomics:GTL and SciDAC Programs (DE-FG02-04ER25627) and NSF award IIS-0946598.

References

ALVAREZ, G. A., KONKLE, T., AND OLIVA, A. 2007. Searching in dynamic displays: Effects of configural predictability and spatiotemporal continuity. *J. Vis. 7*, 14 (12), 1–12.

BALAS, B., NAKANO, L., AND ROSENHOLTZ, R. 2009. A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis. 9*, 12 (11), 1–18.

PROCTER, J. B., THOMPSON, J., LETUNIC, I., CREEVEY, C., JOSSINET, F., AND BARTON, G. 2010. Visualization of multiple alignments, phylogenies and gene family evolution. *Nature Methods 7*, 3 (Mar.), S16–S25.

ROSENHOLTZ, R., LI, Y., AND NAKANO, L. 2007. Measuring visual clutter. *J Vis 7*, 2, 17.1–1722.