Perceptual Principles for Scalable Sequence Alignment Visualization

Danielle Albers
University of Wisconsin-Madison

Michael Gleicher
University of Wisconsin-Madison

Problem

Goals:

Construct a large-scale sequence alignment visualization tool

Use perceptual principles to understand where current tools fail to scale
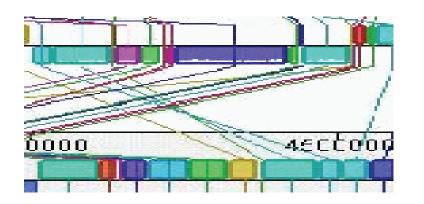
Motivation: The quantity and complexity of genomics data is rapidly growing beyond what current sequence alignment visualization tools can handle
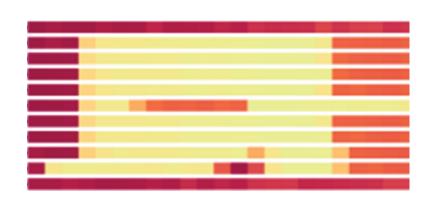
Tools must handle longer genomes, more genomes, many-to-many correspondences, probabilistic information, and other complex data. Scalability in current tools is often hindered because the visual design breaks down as the interpretation problems scale up (see Mauve₁ in Figure 1). By understanding the perceptual limits that cause this breakdown, we hope to design systems that can push them. We survey recent perceptual literature and apply the principles to Mauve, a successful alignment visualization tool, to understand where it fails to scale. We also introduce early efforts at using these principles to construct a new large-scale overview visualization tool (Figure 2).

Pre-Attentive Phenomena

Principle 1: Encodings can allow a user to rapidly identify targets in cluttered environments

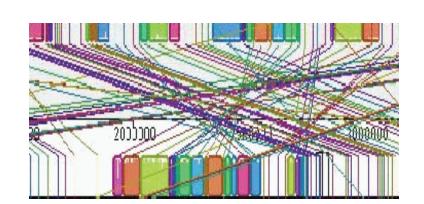**Application:** Simplify visual search by making certain groups of objects "pop out"

The advantages of designing with pre-attentive phenomena in mind are well-understood in the visualization community. However, care must be taken to use pre-attentive mechanisms effectively and avoid unwanted pop-out and similar pre-attentive effects. Mauve (left) employs a color scheme that causes pre-attentive association of unrelated regions. Our design (right) tries to avoid this pitfall: large fields of colors can be matched and texture patterns suggest sequence events.

Pre-Search Processing

Principle 2: Pre-attentive features and structure develop the context of the scene

Application: Blending structural scaffolding with feature data allows more rapid identification of regions of interest and more efficient search

Pre-search processing initiates the visual search process by collecting information to guide search. Alignment visualization can better facilitate pre-search processing by encoding significant details in the low-resolution properties of visual encodings. Mauve's tangled synteny lines (left) cannot be effectively interpreted pre-attentively, whereas large regions of similarity can be determined pre-attentively by color in our design (right).
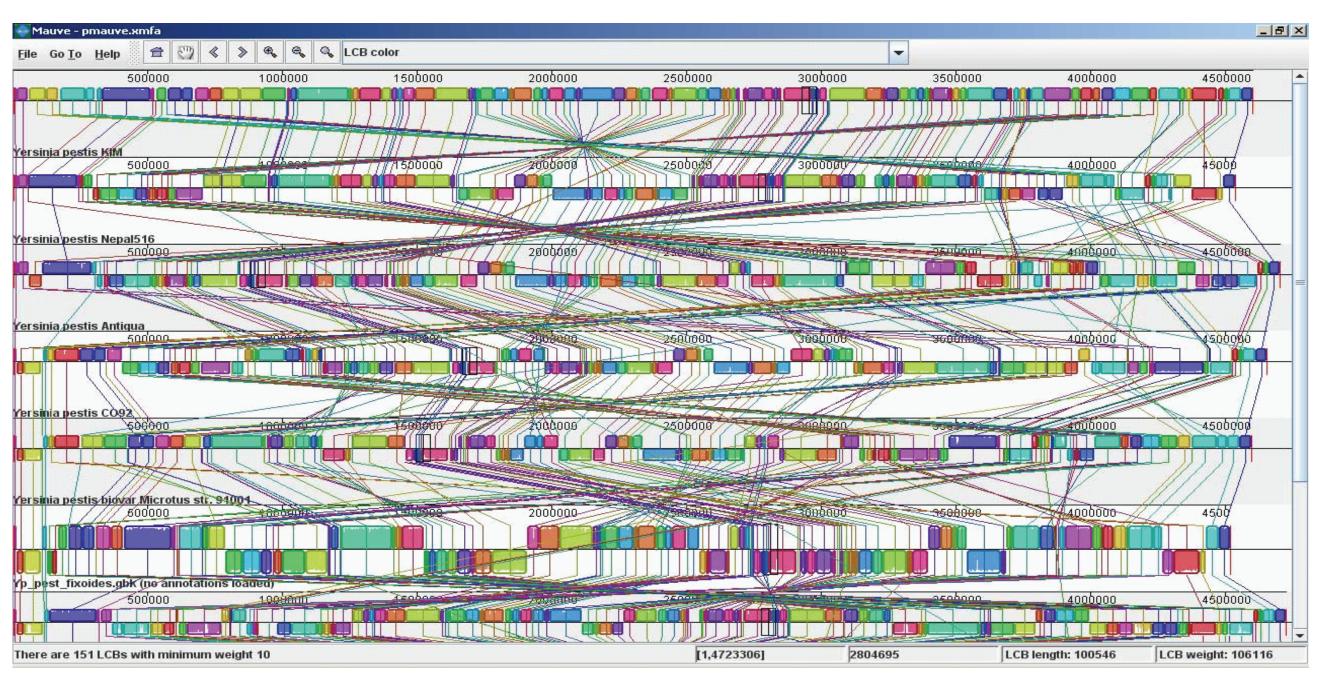

Figure 1: Mauve is a successful tool for visualizing alignments among a small number of sequences, but breaks down as the number of sequences or the complexity of the relationships grow. Even the tool's authors feel it breaks down around seven genomes. Much of the scalability concerns can be explained by perceptual principles.
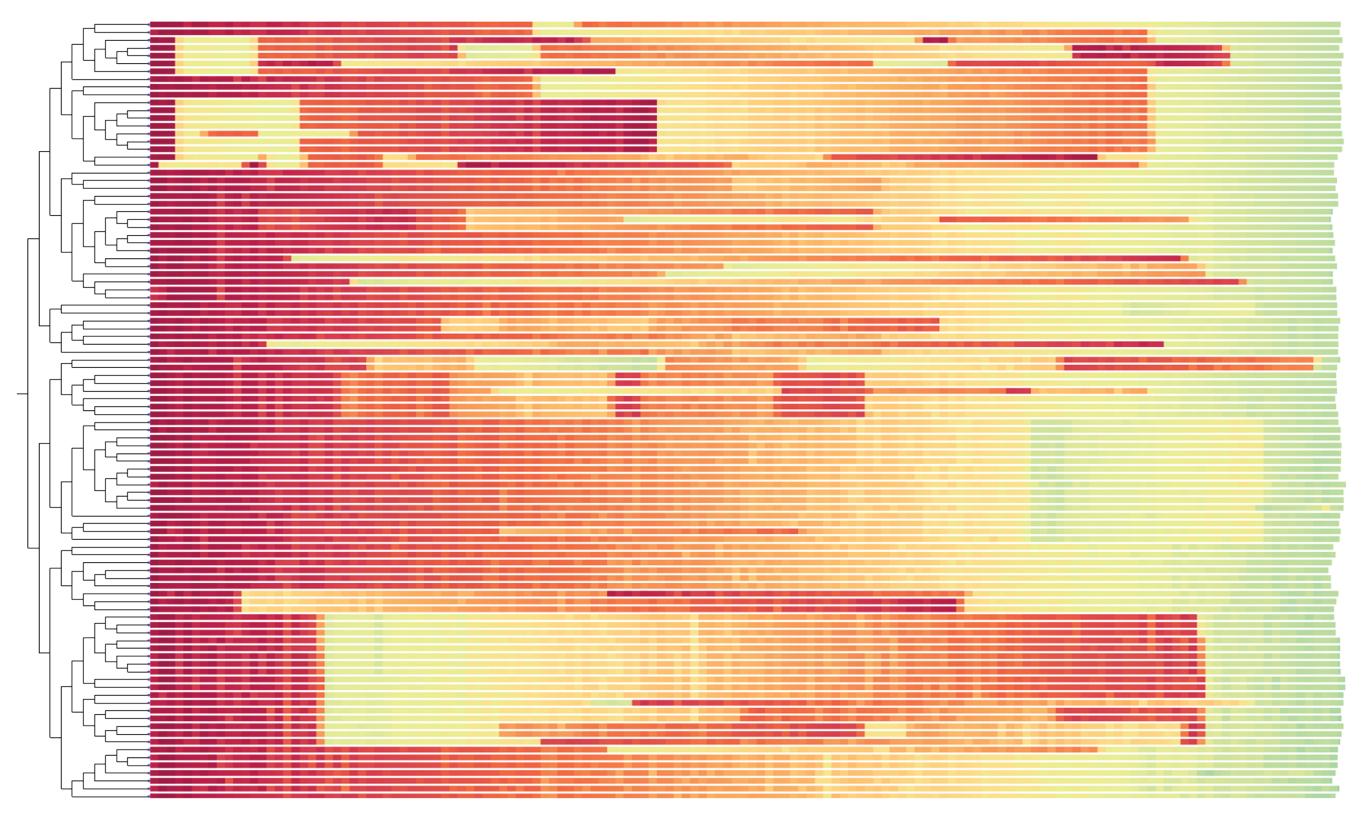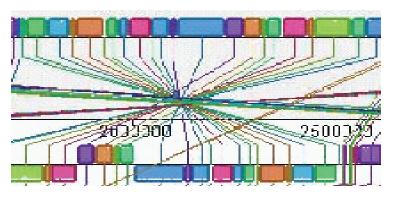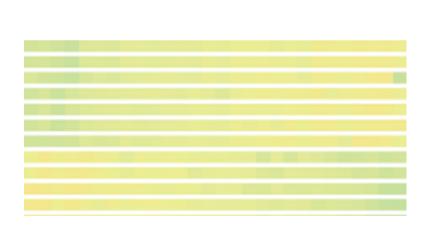

Figure 2: A prototype visualization designed based on the principles presented in this survey. The visualization shows 100 genomes, each comprised of approximately 5,000 genes. The genes are colored by orthology and sorted by local ordering. The genomes are arranged according to their evolutionary tree (left).

Visual Search

Principle 3: Users must scan their attention over the scene to search for targets not found pre-attentively

Application: Designing structure to aid search reduces cognitive cost and improves data processing
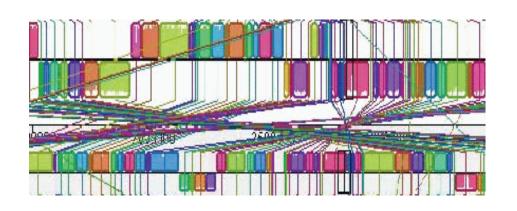

Without perceptual aid, search tasks can be cognitively expensive and time-consuming. Users are generally comfortable searching displays using traditional reading orders. Synteny lines impose a non-linear reading order (left), forcing the user to follow lines to search the data. The surveyed findings suggest a conventional reading order allows the user to methodically scan the data in a logical ordering and reduces the cognitive cost of visual search.

Visual Clutter

Principle 4: Item quantity, encoding, or layout can bog down cognitive processes and slow visual search

Application: Controlling semantic granularity reduces visual clutter without removing data

Synteny and parallel views frequently fall victim to visual clutter at medium to large scales as ribboning techniques often create tangled webs of synteny lines (left). Our design (right) attempts to avoid clutter except when there is a high density of sequence events. The clutter effectively becomes a texture that signifies many rearrangements; it does not obscure other regions.

Summarization

Principle 5: Statistical summaries provide information about non-attended regions of the visual field

Application: Effective summaries rapidly provide overview information without requiring user attention

Alignment visualization can take advantage of perceptual summarization by using visual encodings that remain meaningful when visually summarized. When blurred (which is what summarization effectively does), a Mauve view becomes a gray mass (left), while the prototype's view tries to retain useful features, such as gradients (right). By employing a perceptually regular color scheme and structure, designers can take advantage of summarization mechanisms.

Future Work

Goals:

Develop a practical tool that provides large-scale alignment visualization

Develop a general understanding of the perceptual principles and how they inform large-scale visualization design

We aim to increase the overall scalability of the tool using the perceptual principles outlined in this work. Furthermore, we will try to increase the generality of the tool to accommodate different complexities of modern genomics data.

References
Genome Evolution Laboratory. *Mauve: Multiple Genome Alignment*. http://asap.ahabs.wisc.edu/mauve/, 2010.