Meyer, Munzner, and Pfister have presented a taxonomy visualization system to explore the synteny relationships between the genomes of two fish, the stickleback, and the pufferfish. These relationships are viewable at three levels: genome, chromosome, and synteny block. The visualization tool designed, the Mizbee browser, allows interactivity with the data, providing side by side views of the relationship types across all three ranges. After its design, the Mizbee browser was utilized in two case studies to enhance previous analytical work done. Mizbee’s function is to enable researchers to visualize the relationships of conservation between genomic features in two different genomes, or to infer comparisons within the same genome. Synteny relationships are generally described as on the same chromosome, but because of the numerous number of features contained within, syntenic blocks have been created that represent contiguous sets of features. One problem with these blocks is that despite the numerous algorithms that have been produced, they require so many parameters to utilize that uncertainty and false positives are rampant amongst any results. The purpose of the Mizbee browser is an attempt to show these different relationships at different scales, demonstrated as comprehensible visual relationships. The Mizbee browser is the result of a three-step design study. The first step involved a detailed characterization of the domain questions. The second step involved a taxonomic analysis of the visual encodings that would be suitable for the visualization of this data. The third and final step was the production of the Mizbee browser in the Processing programming language.

In the Data Characterization stage, the raw data was obtained from two different biologists who utilize syntenic datasets as part of their analytic process. The syntenic blocks were diversified into three main layers. The highest level is the genome, which contains a list of chromosomes, the second level. In the chromosome level, a list of blocks is contained, the third level. In the third level, features such as chromosome id, coordinate in the sequence, length, orientation, and many others are recorded. Some design challenges during this stage included: the vastly varied length of nucleotides (less than a dozen to over one billion). A base set of 14 domain questions were compiled during this process, while question 14 “What are the differences between individual nucleotides of feature pairs” was deemed not as important, as many other visualization tools allowed this analysis, while the other 13 questions were unique to Mizbee. The domain questions and feature sets developed during this stage were then utilized in stage two.

In the Visual Encoding stage, the domain questions produced in step one were devolved into a taxonomy that could then be used to generate the vis idiom. Analysis of this data resulted in the design choice to represent connections in the blocks in the forms of lines, connecting matching segments. At the chromosome level, one to many relationships occur, so mere line connections would provide too much visual clutter. Color was decided on to demonstrate the relationships. At the highest level, the genome, the one to many relationships become complex, many to many, relationships. A continuous vs discrete taxonomy layout scheme was decided upon, as color and lines, even curved ribbons, were too inconclusive. Unique lines with minimal crossings, no obscuring of lines by segments, and minimal variance in length were achieved during this process. A circular layout was decided upon for the genome level to reduce the amount of variation in the length of the curves.

In the final stage, the Mizbee browser was born. The data levels and domain questions from stage one as well as the design decisions made in stage two resulted in the design of the Mizbee. The final design of the Mizbee was an iterative, collaborative process with the two biologists who seemed to use the visualization to aid in their own case studies. In the Genome view, the colormap that was decided upon uses an eight-element legend from ColorBrewer. Genomes with more than eight chromosomes, the colors repeat, as representing each as a unique color would become too noisy and visually distracting. Edge bundling greatly assisted in the reduction of visual clutter, as well as filtering the rendered blocks. In the Chromosome view, answers about proximity, location and size are demonstrated. More precise spatial relationships can be determined from this view. Finally, in the Block View, the most detailed data can be viewed. The features are demonstrated as ‘glyphs’ to indicate their orientation. Color encoding is uniform throughout.