Title

Volume visualization of multiple alignment of genomic DNA

Permalink

https://escholarship.org/uc/item/9v25c6jb

Authors

Shah, Nameeta
Weber, Gunther H.
Dillard, Scott E.
et al.

Publication Date

2004-05-01

Volume Visualization of Multiple Alignment of Genomic DNA

Nameeta Shah[1,2,*] Gunther H. Weber[1,2,*] Scott E. Dillard[1,*] Bernd Hamann[1,2,*]

[1] Center for Image Processing and Integrated Computing (CIPIC), Department of Computer Science,
One Shields Avenue, University of California, Davis, CA 95616-8562, U.S.A.
[2] Visualization Group, National Energy Research Scientific Computing Center (NERSC),
Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, U.S.A.

Abstract

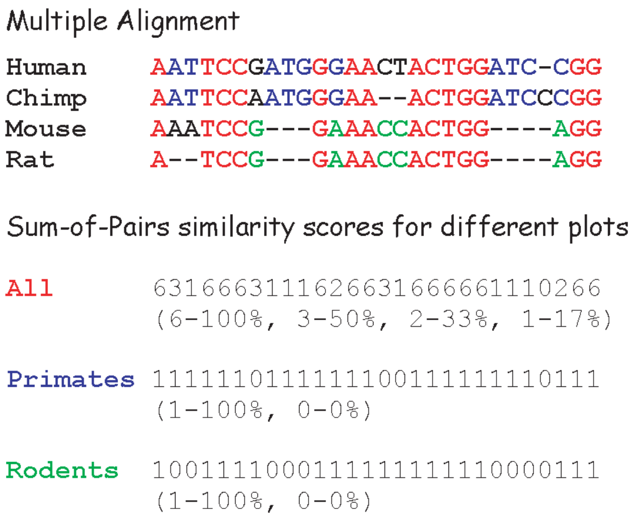
Genomes of hundreds of species have been sequenced to date and many more are being sequenced. As more and more sequence data sets become available, and as the challenge of comparing these massive "billion basepair DNA sequences" becomes substantial, so does the need for more powerful tools supporting the exploration of these data sets. Similarity score data used to compare aligned DNA sequences is inherently one-dimensional. One-dimensional (1D) representations of these data sets do not effectively utilize screen real estate. We present a technique to arrange 1D data in 3D space to allow us to apply state-of-the-art interactive volume visualization techniques for data exploration. We provide results for aligned DNA sequence data and compare it with traditional 1D line plots. Our technique, coupled with 1D line plots, results in effective multi-resolution visualization of very large aligned sequence data sets.

Keywords: Multiple alignment, Hilbert curve, volume visualization

1 Introduction

1.1 Multiple Alignment

The human genome consists of about three billion basepairs, of which only a small percentage is well understood. In order to decipher the rest of the genome, and to understand general principles of genome structure and function, biologists compare genomes, or parts of genomes, of different species. Alignment is an extensively used method for comparing DNA sequences. Currently, biologists compare genomes of many species at different evolutionary distances [Frazer et al. 2003] by examining multiple alignments. A multiple alignment is a set of sequences in a "rectangular arrangement," where each row consists of one sequence padded by gaps, such that the columns highlight similarity/conservation between positions (http://www.cryst.bbk.ac.uk/BCD/bcdgloss.html). Figure 1 shows an example of a multiple alignment of four sequences, where the characters A, T, C and G represent the bases adenine, thymine, cytosine and guanine, respectively. Below the alignment we show similarity scores for the multiple alignment. The similarity score for each column shows the level of conservation among sequences, considering all possible pairwise comparisons of characters (six in the case of four species). Different schemes are used to calculate similarity scores, including *entropy, sum-of-pairs, weighted sum-of-pairs, parsimony*(http://lepo.it.da.ut.ee/ mremm/kurs/multali.htm) etc. We assume that similarity scores are provided as input for our visualization purposes.



```
Multiple Alignment

Human      AATTCCGATGGGAACTACTGGATC-CGG
Chimp      AATTCCAATGGGAA--ACTGGATCCCGG
Mouse      AAATCCG---GAAACCACTGG----AGG
Rat        A--TCCG---GAAACCACTGG----AGG

Sum-of-Pairs similarity scores for different plots

All        6316663111626631666661110266
           (6-100%, 3-50%, 2-33%, 1-17%)

Primates   1111110111111100111111110111
           (1-100%, 0-0%)

Rodents    1001111000111111111110000111
           (1-100%, 0-0%)
```

Figure 1: Multiple Alignment of four species: human, chimp, mouse and rat.

1.2 Current Visualization Tools

Several tools for visualization of alignment data are publicly available. One highly popular and successful tool is VISTA [Mayor et al. 2000]. VISTA represents the level of conservation between species as a curve calculated by sliding a window of predefined size over the given alignment and computing the average similarity score over the window. VISTA shows pairwise similarity scores. Phylo-VISTA [Shah et al. 2004] extends the VISTA concept to the visualization of multiple(more than two) alignments. Other commonly used tools like MultiPipMaker [Schwartz et al. 2003] and SynPlot [Göttgens et al. 2001] also use 1D line or dot plots.

1.3 Motivation

With advancements in sequencing technology, increases in computational power, and the development of better computational methods, it is now possible to align several million-basepair-long sequences. It is clear that the need exists, or will exist in the very near future to develop new visualization techniques to support the interactive, visual exploration of billion basepair-long sequence data.

[*] {nyshah, ghweber, sedillard, bhamman}@ucdavis.edu

Earlier work by Wong *et al.* [Wong et al. 2003] used space-filling *Hilbert curves* to transform sequential data into 2D space. This transformation allows one to display one million basepairs using a 1000×1000 pixel image. Application of digital image processing filters to such images reveals interesting patterns in the data. This work motivated us to represent multiple alignment data in 3D space, embedded in a fixed volume by using 3D space-filling curves. This approach allows us to apply various volume visualization techniques to render the data. We use hardware-accelerated volume rendering for visualization. Often, choosing a transfer function for volume rendering is not a trivial task. In our application, however, we specify a transfer function based on parameters relevant from the perspective of the driving biological problem.

2 Our Approach

2.1 From Sequential, 1D, Data to Volume, 3D, Data

A naïve approach for arranging sequential data in 3D space is using a scan line traversal in 2D planes and stacking these planes in perpendicular direction. In a 64^3 volume grid, for example, such an arrangement will place the 0^{th} position(mapped to (0, 0, 0) in 3D space) and 4096^{th} position(mapped to (0, 0, 1) in 3D space) next to each other. When two positions that are distant in the 1D sequence happen to be adjacent in a 3D arrangement, interpretation of data/visuals is difficult. To mitigate this problem, a 3D arrangement that maximizes spatial coherence should be used. Work by [Keim et al. 1995] and [Voorhies 1991] has shown that a Hilbert curve-based mapping is among the most coherent space filling curves. Coherence is defined as the amount by which neighboring pixels (voxels in our case) are at sequential positions on the curve [Wong et al. 2003]. Figure 4 shows a 3D Hilbert curve. We map a score at position *i* in a multiple alignment to a position in 3D space using the algorithm described in [Max 1998].

2.2 Volume-based Visualization

2.2.1 Color Channels

Consider a multiple alignment of sequences of four species: human, chimp, mouse and rat, see figure 1. Biologists will be interested in:

1. conserved features in all four species,
2. primate-specific features, and
3. rodent-specific features.

These features can be identified by considering the following three different similarity plots and comparing them:

1. a similarity plot for all four species,
2. a similarity plot for human and chimp (primates), and
3. a similarity plot for mouse and rat (rodents).

In this case, we can take advantage of three color channels, red, green and blue, to visualize three similarity plots: We map the first similarity plot to the red channel, second plot to the green channel and third plot to the blue channel. This approach allows a user to compare common and distinct features of all plots. Different colors highlight different features of the data. For example, blue represents a primate-specific feature absent in rodents, green represents a rodent-specific feature absent in primates, and cyan represents a feature specific to both primates and rodents. Further, white (the combination of all colors) represents a region conserved among all species.

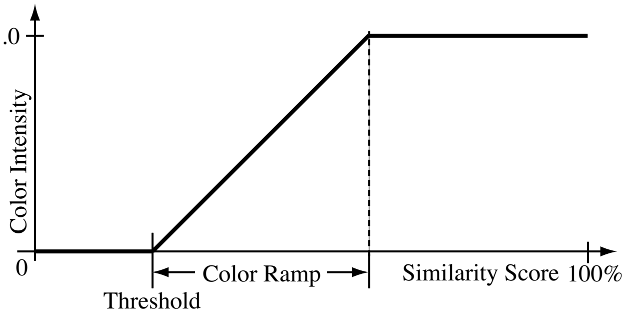


Figure 2: Transfer function.

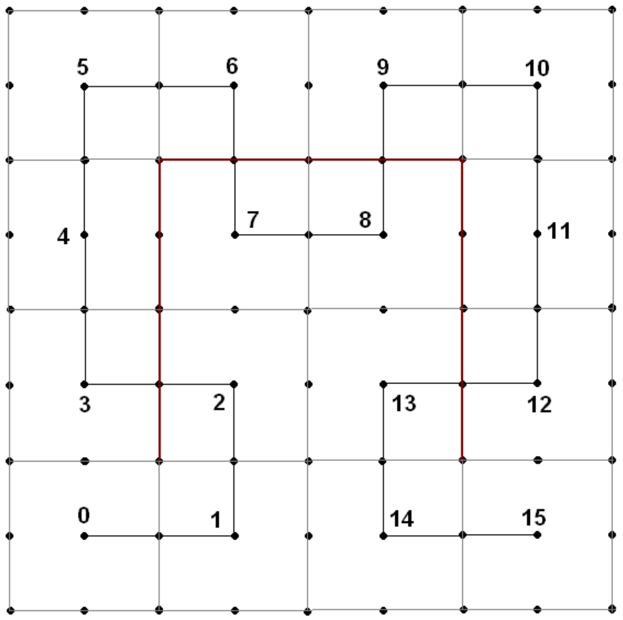


Figure 3: Multiresolution Hilbert curve embedded in 2D space. The black curve is a high-resolution Hilbert curve; a lower resolution version is shown in red .

2.2.2 Transfer function

A separate transfer function is associated with each color channel. We use a linear transfer function like the one shown in Figure 2. We use a user-defined similarity score threshold, below which everything is rendered transparently. The slope of the function is adjustable.

2.3 Annotations

For analysis of multiple alignment data, biologists often need additional biological information, such as information concerning of a known gene model. For example, it is desirable to show the start and end coordinates of a gene, exons (the protein coding part of a

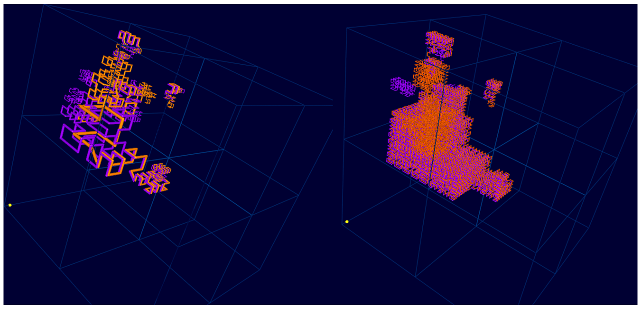


Figure 4: Annotations. Right: annotations drawn using multiresolution Hilbert curve; left: annotations drawn using high resolution Hilbert curve. Purple and orange pipes represent two different sets of annotations.

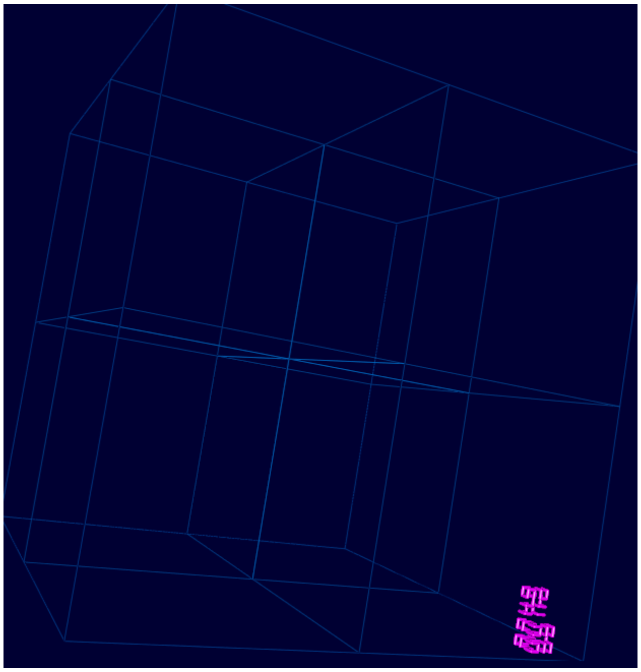


Figure 5: Effective use of screen real estate. Two annotations of sizes 30 and 50 separated by ten basepairs on 262,144 basepair long sequence are distinctly visible.

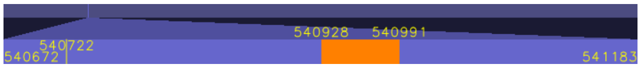


Figure 6: Control for navigation on original 1D data sequence. The upper bar corresponds to the complete sequence, and the lower bar corresponds to the displayed portion of a sequence. A highlighted region in the upper bar represents the displayed portion of the complete sequence and is connected to the lower bar, showing the correspondence of the displayed sequence portion to the complete sequence. The highlighted region in the lower bar denotes the selected octant.

gene) etc. This information can be provided in the form of annotations next to genome sequences. Thus, in addition to displaying similarity plots of a multiple alignment, displaying annotations is a major aspect of genomic data visualization. We draw annotations as “pipes” following the 3D Hilbert curve (Figure 4). As the size of an annotation grows larger so does the number of pipes we draw. This method may clutter the display and slow down the interaction. We handle this problem by using a multiresolution Hilbert curve. Figure 3 shows a Hilbert curve drawn in 2D space using two resolutions. The black curve is a high-resolution curve with 15 line segments ; the red curve is a lower-resolution version with just three segments. The image on the right-hand side in Figure 4 shows annotations drawn using a multiresolution Hilbert curve approach, and the image on the left-hand side uses the same annotations drawn as a high-resolution Hilbert curve. We draw lower-resolution curves as pipes with large diameters to show that they cover a greater volume. We plot overlapping annotations by drawing multiple Hilbert curves with a slight offset (Figure 4). As a result of our volume-based visualization approach annotations can also be displayed at a higher resolution. Consider a 262,144 basepair long genome sequence with two annotations with start and end indices of 160 and 190, and 200 and 250. These annotations, when displayed using a traditional 1D plot at 1000-pixel resolution, would occupy the same pixel position on the screen, and the two annotations would be indistinguishable. Using our 3D visualization method the two annotations can be seen distinctly (Figure 5).

3 Implementation

3.1 Volume Renderer

We render images using view-perpendicular slices with a back-to-front ordering. This approach allows us to incorporate transparency into the transfer function. We use a fragment program to evaluate the transfer function [Beretta et al. 2003]. This program compares the value at each voxel with the three thresholds (for the three color channels.) If a value is above the threshold, that color channel is activated for that voxel. If no value is above the threshold, the voxel is transparent rather than black. “Color Ramp”(see figure 2) can be increased to yield a smooth transition between states. In this case, values near the threshold will be only slightly visible while values much greater than the threshold will clearly stand out.

3.2 User Interface

3.2.1 Annotation Controls

An arbitrary number of annotations can be loaded and displayed with our prototype system. User-defined colors are associated with each set of annotations. In addition, the diameters of the pipes can be adjusted.

3.2.2 Navigation Controls

A Hilbert curve is *fractal* in nature [Sagan 1994]. As a result, data string embedded in a 3D volume can be organized using an octree-like structure. For navigation purposes, a currently displayed sequence portion, which always has the shape of a cube, is divided into octants. A user can then select an octant for zooming in. To provide a context for the currently displayed sequence portion, the bounding box of the volume corresponding to the complete sequence can be displayed.

A 3D representation of inherently 1D data poses problems for navigation in 3D space. A user must know the 1D position of (sequence) data, but one can lose one's orientation when navigating in 3D space. We tackle this issue by using a 1D representation of the sequence shown next to the volume display. This representation allows a biologist to keep track of the position within the sequence in a more traditional fashion. This additional display consists of two bars, see Figure 6, where the upper bar corresponds to the complete sequence and the lower bar corresponds to the displayed portion of a sequence. A highlighted region in the upper bar represents the displayed portion of the complete sequence and is connected to the lower bar, indicating the correspondence of displayed sequence portion to the complete sequence. Selecting an octant to zoom in can be accomplished by clicking on the corresponding octant in the volumetric image, or by clicking on the corresponding part of the 1D bar representing the displayed portion of the sequence. Zooming out leads a user to the next lower level of detail. The 1D sequence representation can also be used to move a marker through the 3D display volume. This marker is symbolized by a vertical line in the 1D sequence display. Dragging this line by using a mouse moves the marker through the volume along the Hilbert curve.

4 Results

We have applied our method to multiple alignment datasets that were created using MLAGAN [Brudno et al. 2003]. We have used these two test datasets:

1. Stem Cell Leukemia (SCL) dataset
   The SCL dataset is a multiple alignment DNA sequence data set of sequences from five species: human, mouse, chicken, pufferfish and zebrafish. All sequences contain the SCL gene. The alignment consists of $150,000$ basepairs. These sequences were aligned in order to discover regulatory elements of the SCL gene. Regulatory elements are short DNA sequences consisting of $6-12$ basepairs. They are generally found in a region in front of a gene called *promoter*. The underlying assumption is that they will be conserved in evolutionary distant species because regulatory elements are functionally important.

2. Cystic Fibrosis Transmembrane Conductance (CFTR) dataset
   The CFTR dataset is a multiple alignment DNA data set of sequences from 12 species: human, chimp, baboon, cat, dog, cow, pig, mouse, rat, chicken, fugufish and zebrafish. The sequences are from the region containing a gene coding the CFTR regulator, and nine other genes. The alignment is 16 million basepairs long. This alignment can help biologists with the discovery of regulatory elements as well as their identification of subclass-specific features.

We compare volume-based visualizations of these datasets with 1D similarity plots generated using Phylo-VISTA.

Figure 7 shows the visualization of the SCL dataset. We visualize three similarity plots: The similarity plot for all five species, which is mapped to the red channel; the similarity plot for human-mouse-chicken, which is mapped to the green channel; and the similarity plot for the two fish species, which is mapped to the blue channel. The size of the displayed volume is 64^3. As the dimensions of the volume need to be powers of two in our current algorithm, the volume is not completely filled by the sequence. Purple pipes show exons (protein-coding parts of a gene) for the human sequence. Exons of the mouse sequence are shown as orange pipes. In Phylo-VISTA plots, exons are shown as purple bars below the plot. Two bars exist: An upper one showing exons of the human sequence and a bottom one showing exons of the mouse sequence. A box filter with a width of 50 was used to smooth data for all three similarity scores. A threshold of 25% was used for all plots. A 3D rendering (left) and the corresponding 1D Phylo-VISTA similarity plots (right) are shown.

White spots in the volume-rendered image indicate regions of high conservation in all three similarity plots. These spots are seen as peaks in all three corresponding Phylo-VISTA plots. Green spots are conserved regions in the human-mouse-chicken plot that are absent in the fish plot. Similarly, blue spots are fish-specific conserved regions. The top image shows the entire dataset, while the bottom image shows the circled part of the top image zoomed-in. The conserved region seen in these bottom images contains the regulatory elements of the SCL gene [Göttgens et al. 2002]. In order to compare three 1D plots a user has to inspect them by eye and determine whether there are peaks in different plots at the same position. This analysis approach may create problems, especially when one pixel represents the similarity score for more than one column in a multiple alignment. In the case of the volume-rendered image a user needs to consider only the color to compare all three plots at once.

Figure 8 shows a visualization of the CFTR dataset. This figure shows only one plot for the similarity among all 12 species, mapped to the red channel. The size of the volume is 256^3. The sequence data fills 25% of the volume. White pipes show genes, and purple pipes indicate the exons of the human sequence. The data was smoothed using a box filter of width 50. The threshold was 25%. The correspondence between the 3D and 1D plots is illustrated by thick gray lines. The middle image shows the entire dataset. Exons and their conservation are much more clearly and distinctly visible in the 3D plot than in the 1D plot. The red spots in the first half of the dataset indicate conservation among all species. The top and the bottom images show zoomed-in views from different viewpoints of the same circled part of the middle image. The circled part in the top image indicates conservation of the first exon of a gene and the promoter region. The image to its right shows a zoomed-in view. The conserved region seen in this rightmost top image contains regulatory elements of CAV2 gene [Thomas et al. 2003]. Similarly, the bottom images show the conservation of regulatory elements of CAPZA2 gene [Thomas et al. 2003].

Figure 9 shows the visualization of the CFTR dataset. Three different similarity plots are used in the volume-rendered representation. The similarity plot for primates (human, chimp, baboon), artiodactyls (cow, pig), carnivores (cat, dog) and rodents (mouse, rat) is mapped to the red channel. The green channel is used to display the similarity plot for primates, and the blue channel for showing the similarity plot of rodents. The top images show the entire dataset. We use different thresholds for different channels: 50% for red and 75% for the other two channels. The 3D visualization reveals many more features when compared to the 1D plot. Larger primate-specific and rodent-specific regions can be identified from both 3D and 1D plots. Lighter spots that exist mainly in the first half of the data seen in the 3D image indicate conversation among all four classes. This aspect of the dataset is not visible in the 1D plot. The bottom images show a zoomed-in view of the circled part shown in the top images. White spots denote well conserved exons of the MET gene for all four groups. No overall conservation in the intron (noncoding part of a gene) exists, but we can see highly conserved primate-specific and rodent-specific regions in the intron that might be of potential relevance.

5 Conclusions

We have presented a volume-based visualization technique for multiple alignment data. Our results demonstrate that 3D representations and visualizations of data are quite effective and utilize 3D display space effectively. As a result, we can convey information more compactly, especially for billion-basepair sequence data.

Although developed for a particular biological application our method can be applied to other kinds of massive sequential data sets. Other volume-based visualization techniques, like isosurfacing or plane slicing, etc. can also be used when appropriate for a given application.

6 Acknowledgements

This work was supported by the National Science Foundation under contracts ACI 9624034 (CAREER Award), through the Large Scientific and Software Data Set Visualization (LSSDSV) program under contract ACI 9982251, through the National Partnership for Advanced Computational Infrastructure (NPACI) and a large Information Technology Research (ITR) grant; the National Institutes of Health under contract P20 MH60975-06A2, funded by the National Institute of Mental Health and the National Science Foundation; by the Director, Office of Science, U.S. Department of Energy under contract DE?AC03?76SF00098; and the Lawrence Berkeley National Laboratory through a Laboratory Directed Research Development (LDRD) project. We thank Chris Co and Oliver Kreylos for their helpful discussions. We also thank the members of the Visualization and Graphics Research Group at the Center for Image Processing and Integrated Computing (CIPIC) at the University of California, Davis, and the members of the Genome Sciences Department and the NERSC Visualization Group of the Lawrence Berkeley National Laboratory for their support.

References

Beretta, B., Brown, P., Craighead, M., Everitt, C., Hart, E., Leech, J., Licea-Kane, B., Poddar, B., Sandmel, J., Schelter, J. P., Seetharamaiah, A., and Triantos, N., 2003. GL_ARB_fragment_program Specification. Online OpenGL Extension Registry, Aug. Available at `http://oss.sgi.com/projects/ogl-sample/registry/ARB/fragment_program.txt`.

Brudno, M., Do, C., Cooper, G., Kim, M. F., Davydov, E., Green, E. D., Sidow, A., and Batzoglou, S. 2003. Lagan and multi-lagan: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research 13*, 4, 721–731.

Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., and Hardison, R. C. 2003. Cross-species sequence comparisons: A review of methods and available resources. *Genome Research 13*, 1, 1–12.

Göttgens, B., Gilbert, J. G. R., Barton, L. M., Grafham, D., Rogers, J., Bentley, D. R., and Green, A. R. 2001. Long-range comparison of human and mouse scl loci: Localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. *Genome Research 11*, 1, 87–97.

Göttgens, B., Barton, L. M., Chapman, M. A., Sinclair, A. M., Knudsen, B., Grafham, D., Gilbert, J. G., Rogers, J., Bentley, D. R., and Green, A. R. 2002. Transcriptional regulation of the stem cell leukemia gene (scl) comparative analysis of five vertebrate scl loci. *Genome Research 12*, 5, 749–759.

Keim, D. A., Ankerst, M., and Kriegel, H.-P. 1995. Recursive pattern: A technique for visualizing very large amounts of data. In *Proceedings of IEEE Visualization 1995*, IEEE Computer Society, Atlanta, Georgia, G. M. Nielson and D. Silver, Eds., IEEE Visualization, Annual Conference Series, IEEE, 279–288.

Max, N. L. 1998. Visualizing hilbert curves. ieee visualization 1998. In *Proceedings of IEEE Visualization 1998*, IEEE Computer Society and ACM, North Carolina, USA, D. S. Ebert, H. Rushmeier, and H. Hagen, Eds., IEEE Visualization, Annual Conference Series, IEEE, 447–450.

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S., and Dubchak, I. 2000. Vista : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics 16*, 11, 1046–1047.

Sagan, H. 1994. *Space-Filling Curves*. Springer-Verlag.

Schwartz, S., Elnitski, L., Li, M., Weirauch, M., Riemer, C., Smit, A., Green, E. D., Hardison, R. C., Miller, W., and Program, N. C. S. 2003. Multipipmaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Research 31*, 13, 3518–3524.

Shah, N., Couronne, O., Pennacchio, L. A., Brudno, M., Batzoglou, S., Bethel, E. W., Rubin, E. M., Hamann, B., and Dubchak, I. 2004. Phylo-vista: interactive visualization of multiple DNA sequence alignments. *Bioinformatics 20*, 5, 636–643.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., Mcdowell, J. C., Maskeri, B., Hansen, N. F., Schwartz, M. S., Weber, R. J., Kent, W. J., Karolchik, D., Bruen, T. C., Bevan, R., Cutler, D. J., Schwartz, S., Elnitski, L., Idol, J. R., Prasad, A. B., Lee-Lin, S.-Q., Maduro, V. V. B., Summers, T. J., Portnoy, M. E., Dietrich, N. L., Akhter, N., Ayele, K., Benjamin, B., Cariaga, K., Brinkley, C. P., Brooks, S. Y., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S.-L., Huang, M. C., Karlins, E., Laric, P. L., Legaspi, R., Lim, M. J., Maduro, Q. L., Masiello, C. A., Mastrian, S. D., Mccloskey, J. C., Pearson, R., Stantripop, S., Tiongson, E. E., Tran, J. T., Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L. S., Young, A. C., Zhang, L.-H., Osoegawa, K., Zhu, B., Zhao, B., Shu, C. L., Jong, P. J. D., Lawrence, C. E., Smit, A. F., Chakravarti, A., Haussler, D., Green, P., Miller, W., and Green, E. D. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature 424*, 14, 788–793.

Voorhies, D. 1991. Space-filling curves and a measure of coherence. In *Graphics Gems II*, Academic Press, J. Arvo, Ed., Graphics Gems, 26–30.

Wong, P. C., Wong, K. K., Foote, H., and Thomas, J. 2003. Global visualization and alignment of whole bacterial genomes. *IEEE Transactions on Visualization and Computer Graphics 9*, 3, 361–377.

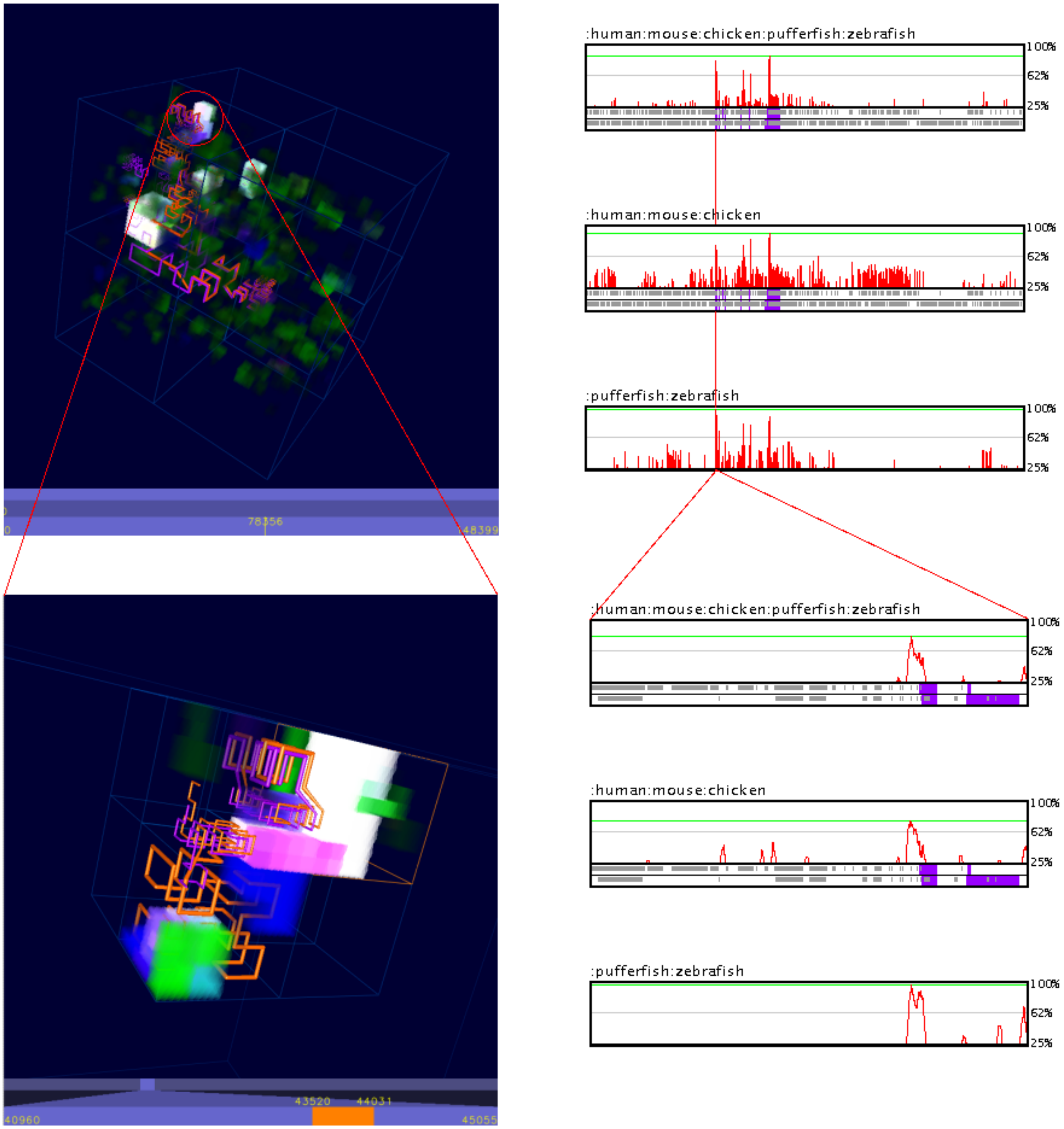


Figure 7: Using volumetric rendering approach to discover a regulatory region for the SCL gene: The white regions in the volume-rendered image correspond to regions that are highly conserved in all three considered similarity scores. The volume-rendered representation allows users to detect these regions instantaneously, without the need to compare multiple plots. Corresponding 1D plots created with Phylo-VISTA.

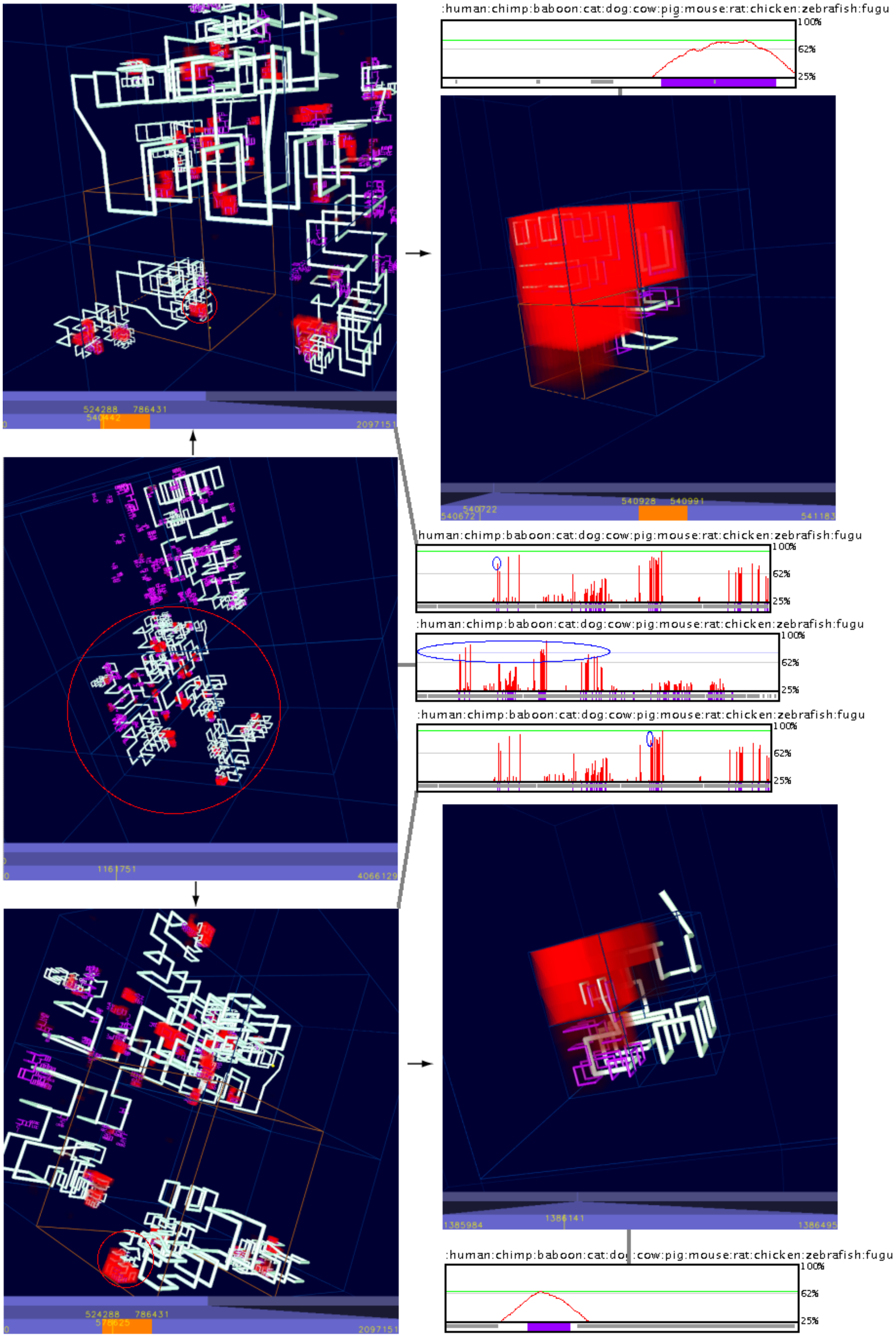

:human:chimp:baboon:cat:dog:cow:pig:mouse:rat:chicken:zebrafish:fugu
100%
62%
25%
524288
786431
540442
0
2097151
540928
540991
540722
540672
541183
:human:chimp:baboon:cat:dog:cow:pig:mouse:rat:chicken:zebrafish:fugu
100%
62%
25%
:human:chimp:baboon:cat:dog:cow:pig:mouse:rat:chicken:zebrafish:fugu
100%
62%
25%
:human:chimp:baboon:cat:dog:cow:pig:mouse:rat:chicken:zebrafish:fugu
100%
62%
25%
1161751
0
4066129
524288
786431
578625
0
2097151
1386141
1385984
1386495
:human:chimp:baboon:cat:dog:cow:pig:mouse:rat:chicken:zebrafish:fugu
100%
62%
25%

Figure 8: Visualizing the CFTR data set: The red spots indicate similarities among all species in a region corresponding to the CAPZA2 gene. In the volume-rendered image, exons and their conservation are much more clearly and distinctly visible than in the 1D plot.

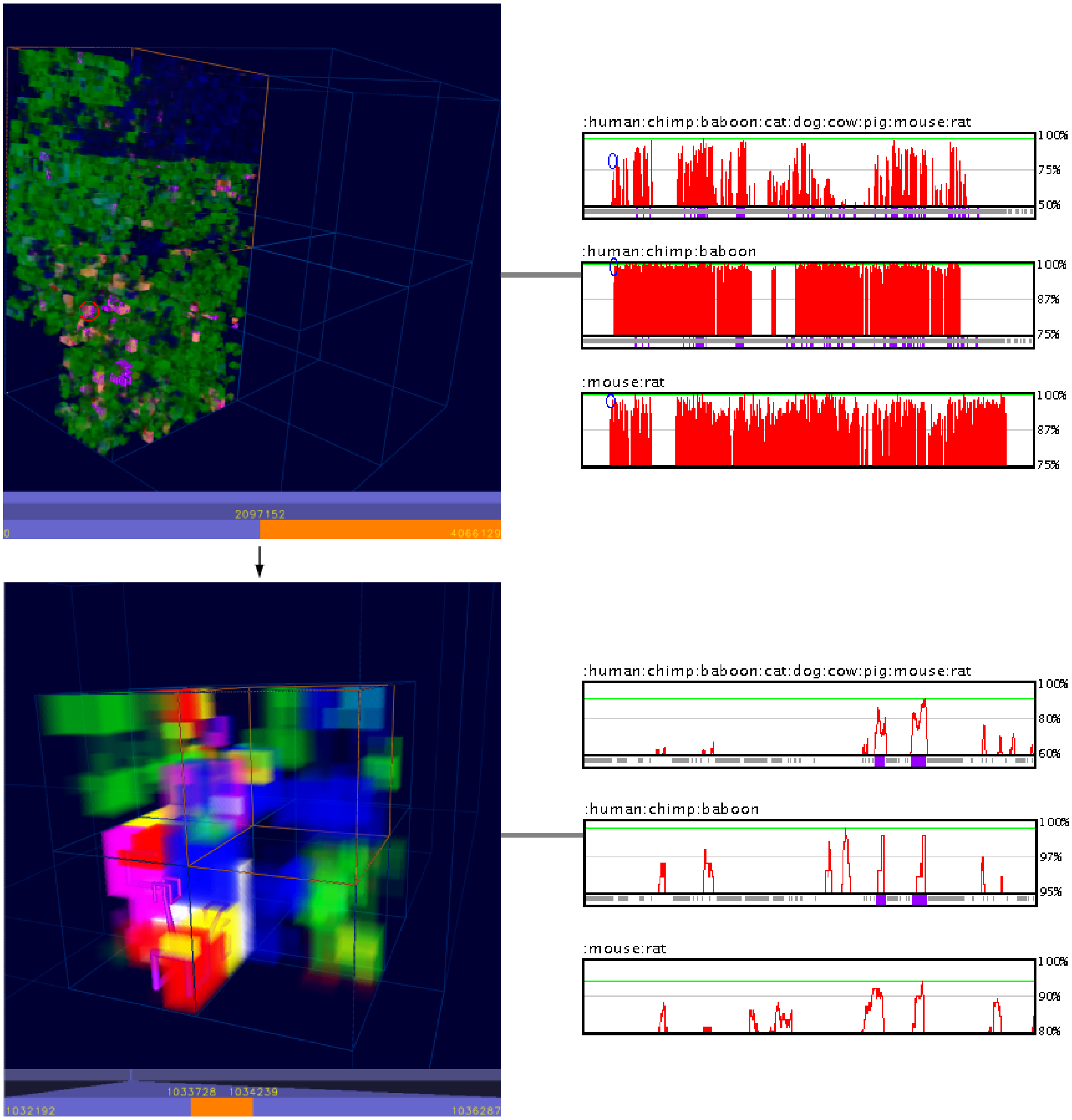


Figure 9: Visualizing three different similarity scores for the CFTR data set. Larger primate-specific and rodent-specific regions can be seen in both 3D and 1D plots. Lighter spots that are visible mainly in the first half of the data in the 3D image indicate conversation among all four classes. This aspect of the dataset is not visible in the 1D plot. The lower images show the well-conserved region of the MET gene.